

## Incompatibility classes and fruit set in natural populations of arctic bramble (*Rubus arcticus* L.) in Finland

JUSSI TAMMISOLA

Department of Plant Breeding, University of Helsinki  
SF-00710 Helsinki, Finland

**Abstract.** In Finnish natural populations, arctic bramble proved uniformly self-incompatible. In vigorous and richly flowering populations, the intensity of fruit set is governed by the number of incompatibility classes present. Most non-fruiting populations contain only one incompatibility class (and most likely only one clone) and therefore totally lack compatible pollen. Richly fruiting populations usually contain at least three incompatibility classes.

A clone with an estimated size of 80 metres and age of 160 years was found. This supports the vegetative burst explanation for the "sudden appearance" of arctic bramble populations. For the purposes of plant breeding, a large genetic collection is required. In cultivation, a thorough mixture of at least three varieties is recommended.

The study was dynamically optimized. A computerized guidance system was constructed, which analyzed the accumulating results and yielded recommendations for forthcoming crosses. For the analysis into equivalence classes of incompatibility, a stepwise clustering algorithm of the single move type, based on the maximum likelihood principle, was introduced. This partitioning was based on the number of seeds in a cross, considering it as a stochastic variable. Seed number probability distributions in intraclass and interclass pollinations were acquired utilizing non-parametric density estimation. Finally, both incompatibility class and seed probability estimates were adjusted together iteratively. A recommendation algorithm was produced, based on a partially heuristic principle of maximum lability — maximum stability. With the aid of the guidance system, incompatibility classes could be resolved using about one tenth of the number of crosses required in a conventional system.

---

Index words: Arctic bramble, *Rubus arcticus*, incompatibility, fruit set, wild berries, population structure, clonal pattern, plant breeding, reproductive ecology, dynamic programming, stepwise clustering, density estimation, iteration, artificial intelligence

---



# Contents

ABSTRACT .....	327
1. INTRODUCTION .....	333
1.1. Characterization of <i>Rubus arcticus</i> .....	333
1.1.1. Taxonomy and distribution .....	333
1.1.2. Aroma .....	333
1.1.3. Habitats .....	335
1.1.4. Spread through growth and reproduction .....	337
1.1.5. Pollination .....	339
1.1.6. Self-sterility .....	341
1.2. Variation of fruit set in nature .....	342
1.3. Hypotheses and study plan .....	344
2. MATERIAL AND METHODS .....	345
2.1. Sampling of populations .....	345
2.2. Sampling of shoots .....	345
2.2.1. Number of ramets to be taken in a population .....	345
2.2.2. Randomization .....	349
2.3. Observations in natural populations .....	350
2.4. Management of the sampled ramets .....	350
2.5. Greenhouse in 1976—78 .....	351
2.6. Experimental field in 1977—80 .....	351
2.7. Techniques in crossing .....	352
2.8. Cytological staining method .....	353
2.9. Statistical test methods .....	353
3. RECURSIVE SYSTEM OF EXPERIMENTS AND ANALYSES .....	354
3.1. Gametophytic incompatibility yields equivalence classes .....	354
3.2. Deterministic model of incompatibility .....	356
3.2.1. Determining equivalence classes via crosses .....	356
3.2.2. Dynamically optimized crossing .....	356
3.2.2.1. Gain in efficiency .....	356
3.3. Stochastic model of incompatibility .....	358
3.3.1. Number of different subdivisions into equivalence classes .....	358
3.3.2. Probability of the actual data on the basis of a hypothesis $H_i$ .....	359
3.3.3. Estimation of the best hypothesis .....	360
3.3.3.1. Likelihood ratio of two arbitrary subdivisions $H_i$ and $H_j$ .....	361
3.3.4. An extended single move method for finding a locally optimal hypothesis .....	361
3.3.4.1. Basic formulas in the algorithm .....	362
3.3.4.2. Extension: amalgamating of two equivalence classes .....	363
3.3.4.3. Comfortable measures .....	363
3.3.5. Choice of the most informative forthcoming crosses .....	364
3.3.5.1. The effect of a single extra cross on likelihood ratio .....	364
3.3.5.2. Preferable crosses .....	365
3.3.5.3. ...and how to find them... ..	367
3.3.5.4. Minimum similarity choice .....	369
3.3.5.5. Maximum lability choice .....	370
3.3.5.6. Choice of the second partner .....	372
3.3.5.7. Allocation of resources between populations .....	373

3.3.6. Estimation of seed number probabilities .....	374
3.3.6.1. Relative seed numbers .....	375
3.3.6.2. A modified Parzen estimation method .....	375
3.3.6.3. Return to a discrete distribution .....	376
3.3.6.4. Estimation of average seed number $E\{\tau_i\}$ in compatible crosses in a population .....	376
3.3.7. Iterative adjusting of equivalence class and seed number probability estimates .....	377
3.3.7.1. Choice between end results of different iterations .....	378
4. RESULTS .....	379
4.1. Grouping of the sampled ramets into incompatibility classes .....	379
4.1.1. Crosses .....	379
4.1.2. No self-fertile clones were found .....	379
4.1.3. Grouping procedure .....	380
4.1.4. Crossing recommendations .....	381
4.1.5. Iterations .....	383
4.1.6. Estimates of probability distributions .....	385
4.1.7. Estimates of equivalence classes .....	387
4.2. Fruiting of the populations in the experimental field .....	389
4.3. Structure and fruit set in natural populations .....	393
4.3.1. Observed quantities .....	393
4.3.2. Population averages according to number of incompatibility classes .....	394
4.3.3. Population averages according to fruit number class .....	397
4.3.4. Population averages according to fruiting zone .....	397
4.3.5. Correlations of population averages .....	400
4.3.6. Quantities within a population according to equivalence class .....	402
4.3.7. Correlations of quantities within a population .....	403
4.3.8. Effect of introducing alien ramets into a population .....	405
5. DISCUSSION .....	408
5.1. Breeding system .....	408
5.2. Dynamically optimized classification .....	408
5.3. In vigorous natural populations, lack of fruit set is generally due to uniclassic constitution .....	409
5.4. Rich fruiting presupposes many incompatibility classes .....	410
5.5. Other factors affecting fruit set .....	411
5.5.1. Equivalence class size .....	411
5.5.2. Density, aggregation and patch form .....	412
5.5.3. Fruiting zone .....	413
5.6. Vegetative or generative burst .....	413
5.7. Plant breeding in natural populations .....	415
6. SUMMARY .....	417
7. CONCLUSIONS .....	420
8. ACKNOWLEDGEMENTS .....	422
9. REFERENCES .....	423
10. SELOSTUS: INKOMPATIBILITEETTILUOKAT JA MARJONTA MESIMARJAN SUOMALAISSA LUONNONESIINTYMISSÄ .....	427
11. APPENDICES .....	430
A. Arctic bramble experimental field at Viikki in 1977—80 .....	430
B. Crossing form .....	431
C. Recording prescriptions, and cross results recorded .....	432
D. Estimated equivalence classes of populations .....	433
E. Crossing recommendations .....	436
F. Details of the crossing technique .....	437
G. Equivalence relation and equivalence classes .....	437
H. Number of crosses needed (in a deterministic, dynamically optimized system) to subdivide a population into equivalence classes of incompatibility .....	437

I.	A single move method for finding a locally optimal hypothesis .....	438
I.1.	One ramet changes the equivalence class .....	438
I.2.	One equivalence class disappears .....	439
I.3.	A new equivalence class appears .....	439
I.4.	Short-hand version .....	439
I.4.1.	Behaviour of the probability part .....	439
I.4.2.	Short-hand version is more favourable .....	441
J.	Average efficiency of procedures I and II in the selection of a proper ramet as the first crossing partner .....	442
K.	Low similarity to its 'own' class is less common in a greater subclass .....	444
L.	Probability of n seeds (taken from a population with three alleles at the incompatibility locus) being of a common incompatibility genotype .....	446



## 1. Introduction

### 1.1. Characterization of *Rubus arcticus*

#### 1.1.1. Taxonomy and distribution

The arctic bramble (*Rubus arcticus* L.) — also called arctic raspberry or 'nectarberry' (in Swedish 'åkerbär', in Finnish 'mesimarja') (TAMMISOLA and RYNNÄNEN 1970) — belongs to the subgenus *Cylactis* (Focke) of the genus *Rubus*. According to FOCKE (1910), *Cylactis* comprises of 14 species; three of them (*R. arcticus* L., *R. saxatilis* L. and *R. humulifolius* C. A. Meyer) have been reported to occur in Finland. The subgenus *Cylactis* was divided into four series: Arctici, Saxatiles, Humulifolii and Xanthocarpi. The series Arctici consists of *R. arcticus* (including ssp. *acaulis* (Michx.) Focke) and *R. stellatus* Sm.

The arctic bramble is a diploid species ( $2n=14$ ). In addition ssp. *acaulis* (on the basis of stomata length, LARSSON 1969), *R. humulifolius* (VAARAMA 1949, 1965) and *R. stellatus* are diploids, whilst stoneberry (*R. saxatilis*) is tetraploid ( $2n=28$ ) (VAARAMA 1939, 1954).

At the present time, also *R. stellatus* Sm. is mostly considered to be a subspecies of *R. arcticus*, ie. ssp. *stellatus* (Sm.) Boiv. emend. Hult. (HULTÉN 1968, 1971, LARSSON 1969, 1980a).

LARSSON (1969) showed that *R. stellatus* crosses freely with *R. arcticus* ssp. *arcticus*, resulting in fully fertile hybrids. HULTÉN (1968) also reported that hybrid swarms are formed between the three subspecies in areas where the ranges overlap. Ssp. *stellatus* is pure only in the Aleutian islands, where the two other subspecies are lacking.

Further evidence for the close relationship between *R. stellatus* and *R. arcticus* ssp. *arcticus* was given by KALLIO (1975a) from biochemical studies of the aroma compounds, and by LARSSON (1969) as well as KOTIMÄKI and HIIRSALMI (1979) from cytogenetic studies.

Since backcrossing is also easy in both directions, (HULTÉN 1968, LARSSON 1969, 1980a, b, TAMMISOLA unpubl., KOTIMÄKI & HIIRSALMI 1979), some new varieties of this newly cultivated species have been bred via hybrids and backcrosses between these two subspecies. For example the 'all-fieldberries' — in Swedish 'allåkerbär' — by LARSSON (1985), in Finnish 'jalomaarain' by HIIRSALMI et al. (1986), are a cross between ssp. *stellatus*, which has provided its desirable robust growth characteristics and ssp. *arcticus*, which has contributed a certain amount of its unsurpassed aroma (LARSSON 1969, 1980b, KALLIO et al. 1980, RYNNÄNEN and DALMAN 1983, HIIRSALMI et al. 1986).

The "genuine" arctic bramble, ssp. *arcticus*, occurs throughout subarctic Eurasia, mainly between 60° and 70° N, but also in Asia in a broader zone, from 50° to 70°. The distribution also extends into the northern parts of North America, ie. Canada and Alaska. Subspe-

cies *acaulis* occurs in the north eastern-most tip of Asia from the shores of the Bering Straits, across Alaska and in subarctic Canada. Ssp. *stellatus* has a narrow distribution range in NW Alaska, the Aleutian Islands and into NE Asia on the shores of the Bering Straits, roughly where the distributions of the other two subspecies coincide (Fig. 1; HULTÉN 1968, 1971, LARSSON 1969).

The three subspecies differ in many characters. Ssp. *acaulis* is tiny and its stem has only one flower, with rather long petals. Ssp. *stellatus* is considerably larger (40 cm) than ssp. *arcticus* (15–25 cm, Figs. 2 and 3), and also has much bigger berries. Its leaves are clearly different in morphology from those of ssp. *arcticus*: they are more roundish, three-lobed but seldom if ever divided into separate leaflets. In addition, it is more resistant to frost, drought and diseases. Having a more 'weedy' growth habit than ssp. *arcticus*, it also competes better with the other species in a cultivated field.

In Finland, only ssp. *arcticus* occurs. In the records from the early decades of this century, the arctic bramble has been reported more or less frequently from every province of Finland. In the southwestern islands, ie. from Åland, it is however almost totally lacking (VAARAMA 1965). Upon reviewing these records, SAASTAMOINEN (1930) concluded that in the western coastal and through the middle parts of Finland there clearly exists an optimal zone, in which the arctic bramble occurs frequently and sets fruit regularly and in abundance (Fig. 5). A more current picture of its occurrence in different parts of Finland, at least in forests and in swampy areas, may be obtained from Fig. 4.

#### 1.1.2. Aroma

The berry of the arctic bramble is used as a foodstuff in several ways. It may be used fresh or frozen, in jam or in liqueur. The best-known product is probably the arctic bramble liqueur (Mesimarja), which is currently so popular that the capacity of the production cannot meet the demand.

The arctic bramble berry is in particular sought after for its unique aroma, which is clearly different from that of other berries (SAASTAMOINEN 1930, LARSSON 1969, 1980b, RYNNÄNEN 1973, KALLIO 1975b). LINNAEUS (1762), who himself succeeded in cultivating the arctic bramble, was well aware of the delicious flavour of its fruits, which he considered as "de smakeligaste af all frukt . . . i hela Europa" [the most delicious of all fruits . . . in whole Europe].

The arctic bramble fruits contain more than two hundred volatile aroma compounds (KALLIO 1976). More than sixty of them have been identified, comprising about

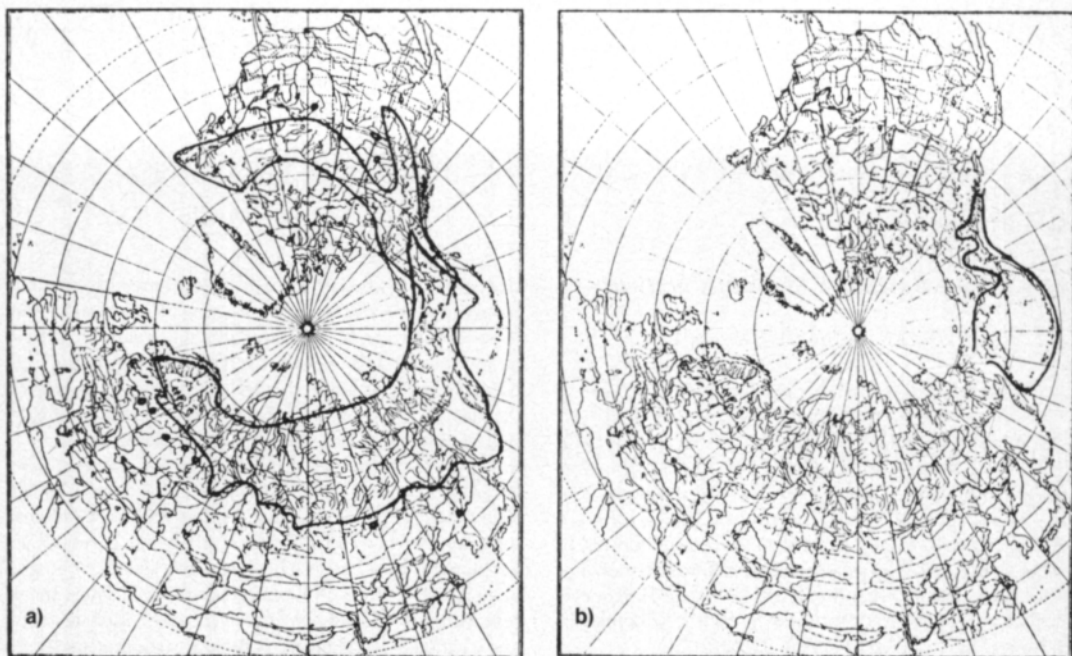


Fig. 1. General distribution of *R. arcticus*.  
 a) *ssp. arcticus* (principally in subarctic Eurasia) and *ssp. acaulis* (in North America).  
 b) *ssp. stellatus* (after HULTÉN 1971).



Fig. 2. Arctic bramble samples growing in pots in a greenhouse at Viikki, at autumn 1976.





Fig. 3. A ramet of a triploid arctic bramble clone (population 116A) at Maivajoki, Kuusamo, 30<sup>th</sup> of July 1976. Notice the leaf morphology: leaflets are more roundish, thick and less lobed than in an ordinary diploid, and often even slightly deformed.

90 per cent of the total aroma concentrate. The most abundant compound among them is 'mesifurane', i.e. 2,5-dimethyl-4-methoxy-2,3-dihydro-3-furanone. In organoleptic analyses, this compound proved to be primarily responsible for the flavour in the press juice of the arctic bramble (KALLIO and HONKANEN 1975). It had been identified earlier in canned alphonso mango. In the measurements of PYYSALO et al. (1977), it expressed extremely high odour value. A closely related compound 'hydroxymesifurane' — i.e. 2,5-dimethyl-4-hydroxy-2,3-dihydro-3-furanone — is also found in the arctic bramble though only in small amounts. This latter compound has been reported earlier in pineapple, beef, strawberry, roasted filberts and almonds, popcorn, and heated maple syrup. Since, however, the threshold value of the latter compound is low, KALLIO (1976) concludes that these two furane derivatives form the basic odour of the arctic bramble.

No research appears to have been done on the aroma constituents in ssp. *acaulis*. According to the information given to her from a Canadian research station, LARSSON (1969) supposed that its fruits lacked the *arcticus*-aroma. VAARAMA (1951), on the basis of information available to him (however not specified), believed that the aroma should greatly resemble that of the arctic bramble. Having tasted the berries in nature, Y. L. A. Mäkinen in a letter comments that they could not be separated from the arctic bramble by aroma.

The 'Alaskan raspberry', ssp. *stellatus*, being tested orally by LARSSON (1969), should however have a 'piquant, smoky but weak' *arcticus*-aroma. This consideration was confirmed by KALLIO (1975a). He found out that the aroma compounds of ssp. *stellatus* are qualitatively very similar to those of ssp. *arcticus*: each of the compounds identified had also been found in ssp. *arcticus*. However, the total amount of volatiles in ssp. *stellatus* is only from one fourth to one third of that in ssp. *arcticus*. And the content of mesifurane (the most important component for the flavour) is from 19 to 44 per cent.

In the hybrid between the two subspecies (ssp. *stellatus* x ssp. *arcticus*), the quantity and quality of the aroma were as a rule intermediate between parents, though the aroma spectrum showed large variation between different samples (KALLIO et al. 1980).

### 1.1.3. Habitats

In Finland the arctic bramble occurs both in natural and in lightly or marginally cultivated habitats. As natural ones may be classified its occurrences on the shores and watersides, as well as in swampy spruce forests and in northern, well-lit groves. Arctic bramble



Fig. 4. Frequency of occurrence of arctic bramble in forests and swampy areas in Finland, according to the third transectional evaluation of Finnish forests (After VAARAMA 1965, modified).

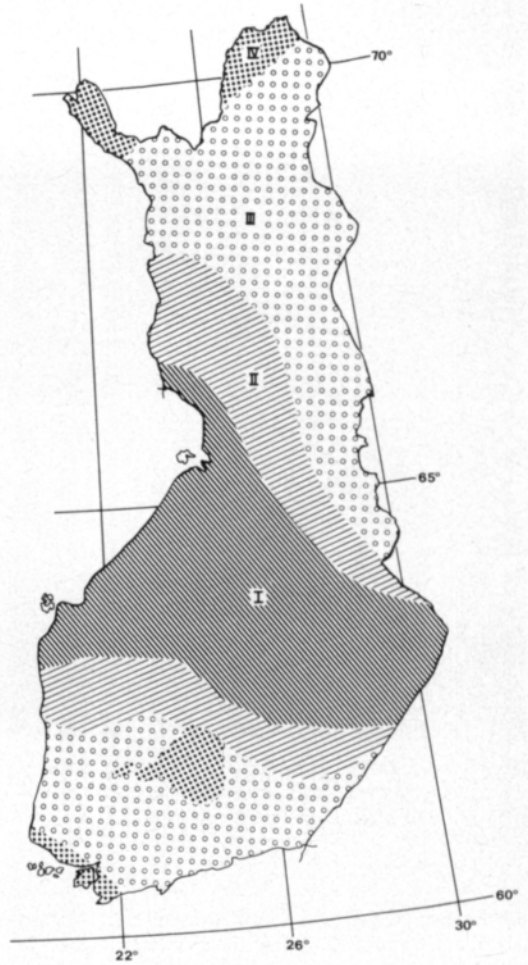


Fig. 5. Frequency of occurrence and fruit set of arctic bramble in Finland, before and up to the year 1930.

I = generally very common and setting fruit richly (optimal zone), II = common and usually setting fruit, III = fairly common but sets fruit only occasionally, IV = rare (After SAASTAMOINEN 1930, modified).

seems to be a weak competitor and will easily be replaced by other species in too crowded or too deeply shadowed plant communities (SAASTAMOINEN 1930, SALMINEN 1948, ERVI et al. 1955; cf. ELVELAND 1983). Hence, its natural habitats are capable of retaining their open nature, without too much competition for light and space exerted by taller plant species.

In N Finland, arctic bramble is commonly encountered along the stony banks of rivers and brooks. There the

vegetation is probably in part controlled (and also occasional spots of bare ground revealed) by ice and floods in spring. (Generative) reproduction in arctic bramble is known to require such spots, since the germinated seeds have been found only on patches of bare ground (SAASTAMOINEN 1930). Reindeers, sheep or other grazing animals may also have a role in keeping the riverside plant communities open enough to permanent growth of *R. arcticus*.

On the sea and lake shores, there often exists a zone of shrubs or low forest, generally dominated by alder species (*Alnus glutinosa* (L.) Gaertn. or *A. incana* (L.) Moench.). On the margins of these zones, vigorous colonies of *R. arcticus* are often to be found. While arctic

bramble shoots are often seen also deeper in the alder zone, they are as a rule weaker there or even have ceased flowering altogether (SAASTAMOINEN 1930, ERVI et al. 1955, TAMMISOLA unpubl.).

The nutritional status of the ground in the arctic bramble stands usually exceeds the level reported from other representative areas and on average, the levels of phosphorus or potassium are higher (ERVI et al. 1955). The high level of potassium may be due to the burning practices (see beneath). The burnt ashes from forest have been used in some instances as a fertilizer for the arctic bramble, with reportedly good results (ERVI et al. 1955). According to VAARAMA (1951), *R. arcticus* grows well on all kind of garden humus, provided it is *not* rich in calcium. SALMINEN (1948) and RYNNÄNEN (1971, 1973) stated that arctic bramble is not restricted by soil pH.

The arctic bramble has considerably benefitted from the ancient, minimal cultural practices of man (LINKOLA 1916). Clearing and burning-over of woodland was exercised for centuries as a regular procedure in agriculture. This provided the arctic bramble with bare ground for the germination of its seeds as well as for vegetative spread (cf. FLINN and WEIN 1977). *R. arcticus* often flourished on these 'kaski' ('slash and burn') fields, which were cultivated for a couple of years and then abandoned or transferred to cattle grazing areas (SAASTAMOINEN 1930). The early cultural practices of man created still favourable environment for the arctic bramble. Occasional ploughing — since it was still minimal cultivation and did not affect the ground too deeply — also inhibited the competition whereby the arctic bramble profited. *R. arcticus* thrived particularly well on the banks of ditches between the field strips. By mowing down his meadows regularly, man kept the vegetation low enough, and prevented bushes or tall herbs from occupying them. Thus the meadows remained rather unchanged for long periods of time. If these meadows were also used for grazing animals, which further controlled the vegetation and — via trampling with their hoofs — created bare patches of land, this type of management also produced a favourable environment for the germination of the seeds. As these meadows age, the arctic bramble diminishes in numbers and in vitality (SAASTAMOINEN 1930).

Since agricultural practices nowadays have altered radically, most of the above mentioned lightly cultivated *habitats* are fast *disappearing*. Due to subsoil drainage, open ditches are rare in the fields. Burning is only seldom used in farming practice today. Ploughing is far more intensive, tilling the ground to a much greater depth, which practically destroys the possibilities of shoot regeneration from root fragments of the arctic bramble.

The effect of modern, heavy ploughing was readily observable when a colleague accidentally ploughed up my first large collection of Finnish *R. arcticus* strains in the autumn of 1974. The next spring only a few miserable shoots could be found in the field.

In forestry, the modern, strict control of forest fires has caused a great decrease in arctic bramble. In former times, forest fires caused by eg. lightning were quite common, and regularly altered the forest ecosystems thus providing the arctic bramble with areas favourable for growth.

Fire has, however, still been used to a certain degree in forest management — in Finland especially during 1925—35 and 1955—65. In the future, controlled burning may be used more, for its preventive effect on *Fomitopsis annosa* (Fr.) Karst., the serious decay causing fungus of Norway spruce (ANON. 1980). Modern forestry management techniques occasionally create suitable conditions for the arctic bramble. Open areas generated by patchwise clear-cutting may give the arctic bramble the possibility to flourish for a period of some years. In addition, the wayside banks of the lumber roads built in the forests often temporarily provide the arctic bramble with suitable habitats.

#### 1.1.4. Spread through growth and reproduction

In this paper, with regard to reproduction, the terminology of HARPER (1977, 1978) has been adopted inasmuch as the term 'vegetative reproduction' has been replaced by 'growth' (see TAMMISOLA 1986).

*R. arcticus* is able to spread vegetatively through its root system. In light soils, the bulk of its roots reach to the depth of 30 cm, and some parts even over 40 cm. In favourable conditions, the rhizomes may extend quite rapidly — in a garden the diameter of a clone often increases by about 0.5 m per year, occasionally even more (SAASTAMOINEN 1930, ERVI et al. 1955, RYNNÄNEN 1973, TAMMISOLA unpubl.). In addition, the roots have the high ability to develop adventitious buds, especially after being injured. Thus, according to old records (SAASTAMOINEN 1930, TAMMISOLA unpubl.), the arctic bramble was able to be fully productive a couple of years following a slight ploughing or burning, and sometimes constituted a continuous under crop beneath eg. a cereal crop which had been previously sown.

In rhizomatous plants, a very great proportion of the biomass may be in the subterranean organs. Figures concerning arctic bramble are, unfortunately, lacking. In *R. chamaemo-*

*rus* L., WALLEN (1986) reports 98 per cent of the biomass to be subterranean.

SAASTAMOINEN (1930) speculates that the sudden "appearance" of dense and richly flowering stands of *R. arcticus* in areas where the forest has been cut down, is based on vegetative spread. The arctic bramble should have been present there already before the clearing, though only as a minor — probably weak in vitality, even sterile — component of the underlayer. ERVI et al. (1955) even speculate that the arctic bramble, and its serious rival *Rumex acetosella* L., may "hide" under ground for several years, being present solely as resting rhizomes. Such a "disappearance" for some years via dormancy in the soil have also been reported or proposed for some other perennial plants (TAMM 1972, TAMMISOLA 1986).

The cutting down of the large trees will abruptly alter the competition between species, especially the competition for light and nutrients, and the arctic bramble, because of its extensive root system, may be able to take advantage of this situation better than most of its rivals. Thus it may even dominate the vegetation for a couple of years, before the species with a stronger competitive ability succeed in taking over.

One should, however, note that minor disturbances of the environment by man, which are advantageous for the vegetative spread of the arctic bramble, may also be conducive for the germination of its seeds. SAASTAMOINEN (1930) noticed that the seeds may germinate successfully only on bare or almost bare ground, which is a characteristic of weak competitors in general (TAMMISOLA 1986). There may be a heavy distribution of *R. arcticus* seeds in the ground where it formerly flourished. Due to its low germination (usually less than 40 %) (ERVI et al. 1955, LARSSON 1957, RYNNÄNEN 1973) and to the thick and hard seed coat, the arctic bramble may have adapted to retain majority of its seeds ungerminated in the ground for long times, waiting for the occasionally suitable conditions for

germination (LARSSON 1955, VAARAMA 1965, TAMMISOLA 1981, 1986).

Hence, without a more detailed inquiry into the stand in question, one can not conclude which of these explanations for sudden appearance is correct. In a pilot study (TAMMISOLA 1981), evidence for the "vegetative burst" explanation was recorded in one stand: there the arctic bramble shoots from both sides of a lumber road were most likely all from the same clone.

Along riverside habitats, the seeds of *R. arcticus* may also be dispersed by spring floods (SAASTAMOINEN 1930). Generally they are however distributed by animals eg. birds. In the North Savo Experimental Station, birds — especially thrush (*Turdus*) species — were frequently attacking the arctic bramble plantations (RYNNÄNEN 1973), while in my experimental fields at Viikki they were hardly at all interested in arctic bramble fruits.

This difference in behaviour may be due to the different abundance of arctic bramble berries in the two districts in question. In North Savo, arctic bramble fruits are still fairly common in nature, while on the South coast, they will be found only occasionally. Predators often concentrate on the prey species which are most common — analogously birds may tend to be disinterested in arctic bramble berries in the districts where as a rule only a few are produced.

In nature, the seeds of *R. arcticus* will not germinate before the next spring. When allowed to dry out, they are known to require a prolonged frost treatment or a treatment with concentrated sulphuric acid before they are able to germinate (SAASTAMOINEN 1930, LARSSON 1957, RYNNÄNEN 1973). However, if the seeds from the very beginning are kept in continuously moist conditions, no frost is needed and the seeds will germinate after seven months of stratification near freezing point (ERVI et al. 1955, LARSSON 1969).

The aerial shoots of the arctic bramble are annual. It overwinters via rhizome buds. These are usually situated near the soil level, just below the litter, and are well developed — thus they grow rapidly in the spring (SAASTAMOINEN 1930, ZELLER 1964).

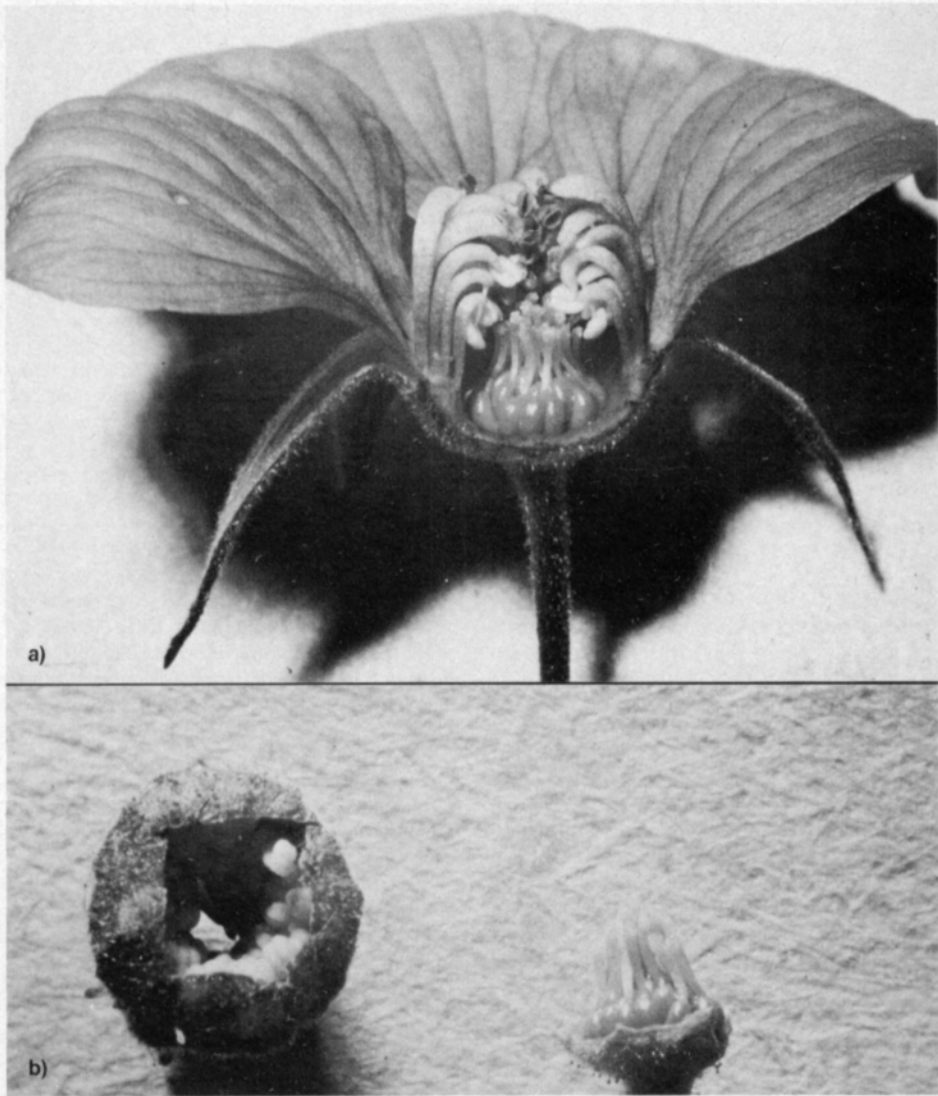


Fig. 6. Flower morphology of *R. arcticus*.

a) In a cross-section of the flower, one can see that the rigid, flattened filaments are in several tight whorls and curved to the centre. Thus, they prevent the entrance to the nectars, and only a few insects, such as bees and bumble bees, or those with a long and thin proboscis, eg. Lepidoptera and *Dolichopus* flies, can penetrate.

b) Being perigynous, the flowers can readily be emasculated by circumcising at the hypanthium with a razor blade. Stamens, sepals and petals will be removed as a common cone, leaving the gynoecium of the bud naked.

#### 1.1.5. Pollination

*R. arcticus* has hermaphroditic flowers and requires insect pollination. Since the filaments of the flowers are thick and curved very closely upon the stigmas, the nectaries are only accessible to insects that are sufficiently strong

(Fig. 6). It has been shown that in experiments with excluders (cages covered with a net with small meshes), small-sized insects are of little importance in the pollination of the arctic bramble (RYYNÄNEN 1973).

In practice, by far the most important pollinators of *R. arcticus* are bumble bees and

honey bees (WARMING 1886, POPPIUS 1903, SILÉN 1906, SAASTAMOINEN 1930, RYNNÄNEN 1973, TERÄS 1985b). Other insects are only occasionally seen on its flowers, eg. butterflies. Various dipteras are also sometimes present, as well as beetles and ants. Generally, though, they have no entrance to the fully concealed nectar, and stigmas, except on damaged flowers. In addition, the pollen-carrying capacity of these insects on their body surface is poor compared to that of bumblebees and honey bees (disregarding even the pollen in the 'baskets', since it has been moistened with nectar and is no more capable of pollinating) (FREE and WILLIAMS 1972, KENDALL and SOLOMON 1973, HIPPA and KOPONEN 1976). The same applies to the visiting speed and activity — especially bumblebees are very active, even at low temperatures and bad weather.

Thrips are often found in the flowers of the arctic bramble, eg. *Thrips major* Uz., *Frankliniella intonsa* Tryb. and *Taeniothrips vulgatissimus* Hal. (RYNNÄNEN 1973). They are, however, small, relatively smooth in body surface and — especially in pre-adult stages — poorly mobile from one flower to another. In addition, the effect of their suction is to ruin the styles and ovaries of the plant, and subsequently also the pollen. (TAMMISOLA unpubl.). Therefore, even if the thrips may sometimes be numerous, their role as pollinators in the arctic bramble can at most be considered negligible. Similar conclusions were drawn by HIPPA and KOPONEN (1976) on cloudberry (*R. chamaemorus* L.).

In *R. arcticus* plots in an orchard in SE Finland (Karila, Mikkeli), TERÄS (1985b) studied food plants and flower visits by bumble bees and cuckoo bumble bees. Arctic bramble was visited by 9 bumble bee species. Of these, three were rare visitors, while six species were regularly found on arctic bramble: *Bombus soroensis* (F.), *B. pascuorum* (Scop.), *B. pratorum* (L.), *B. lucorum* (L.), *B. hypnorum* (L.) and *B. lapidarius* (L.). The role of cuckoo bumble bees (*Psithyrus* species) proved to be negligible as pollinators of *R. arcticus* in the study area. Out of their 1900 flower visits, they only visited arctic bramble once.

In comparison to the other 100 plant species in that study, the flower visits to the arctic bramble were evenly distributed amongst the most frequent bumble bee species present. Noticeably, however, the long-tongued *B.*

*hortorum* (L.) was not interested in the flowers of the arctic bramble, the corolla of which was shorter (4 mm) than that of species from the study area, whose average was 5 mm.

Honey bees may be considered as efficient pollinators, since they are highly flower constant, ie. an individual foraging bee will keep visiting exclusively a single plant species during a foraging trip (eg. PERCIVAL 1947). Hence, pollen of, say, arctic bramble will not be lost ineffectively on the flowers of other species. Since a honey bee is able to inform other bees in the hive when a favourable nectar source has been found (FRISCH 1959), a foraging strategy based on flower constancy is efficient.

In contrast the bumble bee forager — lacking information from other bumble bee individuals — must settle on an individual foraging strategy, based on her own samplings of the available flower resources. This sampling is time consuming but bumble bees have the advantage that they are able to distinguish the already emptied flowers — devoid of nectar — due to their scent, while still in flight. In addition, individual bumble bees follow their own foraging paths or areas, which further diminishes the time during sampling (HEINRICH 1976, 1979, OSTER and HEINRICH 1976, TERÄS 1976, PEKKARINEN and TERÄS 1977, NOUSIAINEN et al. 1978).

An individual bumble bee has been recorded to visit in from one to several plant species on a foraging trip. One of the species, the 'major' one, is usually visited by the bee much more frequently than the other, 'minor' ones. A queen during her lifetime may switch her major species several times, often according to the changes in the abundance of flowering species. In addition, individual bumble bees can have distinct multiple specializations, often unrelated to that species' overall preference (HEINRICH 1976, 1979). Regarding the information available to bumble bees, such a 'majoring' strategy has been shown to be always better than random foraging. The continued small scale sampling ('minoring') is a necessary compromise required to keep track of the resources as they change with time (OSTER and HEINRICH 1976).

The "effectual" flower constancy in bumble bees is higher than one would expect, remembering that usually several species are visited during a single foraging trip. Namely, major proportion of the pollen collected by the bumble bee will be from the majored plant species (PEKKARINEN and TERÄS 1977). Thus, bumble bees may be considered efficient pollinators with regard to their major plant species. Furthermore, the foraging trips are "piecewise flower constant", ie. they usually consist of a fairly continuous sequence of visits on a single plant species, with occasional switching to a sequence on another species (HEINRICH 1976, 1979). Hence, pollination of a minor plant species may also be more efficient than expected.

Unfortunately, however, the results presented by TERÄS (1985b) do not reveal, whether arctic bramble was majored by any individual bumble bees at any time. The conclusion of the study was that *R. arcticus* was not amongst the four most often visited plant species by any *Bombus* species, except for *B.*

*soroensis*. Visits to the arctic bramble by bumble bees of this species, only constituted 8 per cent or less of their total overall flower visits. This small percentage points to the arctic bramble as being a minored plant. Taking into consideration the possible specialisation of individual bees within a bumble bee species, then the bulk of the visits to the arctic bramble may still be exercised by bumble bee individuals majoring the arctic bramble, at least during its principal flowering period in June. This detail may well prove essential considering the pollination efficiency and fruit set in arctic bramble.

#### 1.1.6. Self-sterility

Self-sterility has been reported in many species from various subgenera of the genus *Rubus*, eg. in *R. odoratus* L. (subgenus *Anoplobatus*), *R. allegheniensis* Porter (subgenus *Eubatus*), *R. parvifolius* L. and sometimes in *R. idaeus* L. (subgenus *Idaeobatus*) (EAST 1940, FRYXELL 1957, KNIGHT and KEEP 1962). Although self-sterility predominates in wild *Idaeobati*, it is of little importance within commercial *Idaeobati*, appearing only occasionally in certain progenies of the red raspberry and in hybrids between the red and black raspberries (KNIGHT and KEEP 1962, KEEP 1972, REDALEN 1976). Thus, during its centuries of cultivation (at least since 1548, see VAARAMA 1965), due to a more or less unconscious selection, *raspberry* has turned into a self-fertile species — as has happened with many other cultivated species as well (eg. SCHWANITZ 1967, SIMMONDS 1979).

SAASTAMOINEN (1930) concluded that the flower morphology of *R. arcticus* suggests self-pollination: the filaments are curved to the centre and the anthers of the innermost stamens almost touch the stigmas (see Fig. 6). Such a self-pollination mechanism has been reported to be functioning in some self-fertile *Rubus* cultivars, as well as in the pseudogamously apomictic blackberries (subgenus *Eubatus*) (NYBOM 1985, 1986).

SALMINEN (1948) reported that arctic bram-

ble would have self fertile as well as self sterile strains, the latter however being more common. According to him, the self fertile strains should even be morphologically clearly identifiable from the self fertile ones (ERVİ et al. 1955). RYNNÄNEN (1973), however, criticized the methods applied in these experiments, regarding clonal contamination as very likely.

Isolation experiments on separate clones by LARSSON (1969) suggested that at least all the studied clones of the Arctici-series were self-sterile. This was largely confirmed in pollination experiments by A. Rynnänen on 16 Finnish strains of the arctic bramble (TAMMISOLA and RYNNÄNEN 1970). Twelve of these strains proved clearly self-sterile, in one strain the overall fertility was poor, and in three strains occasional cases of moderate drupelet formation were recorded amongst unequivocally self-sterile reactions.

Such an occasional drupelet formation could be explained by facultative apomixis, or supposing that the strength of the self-sterility is much influenced by the environment. Apomixis, however, has been stated not to occur in diploid *Rubus* species (KNIGHT and KEEP 1962), though it is a common phenomenon among polyploid *Eubati*. In a cytological study with *R. saxatilis* — a tetraploid species in the subgenus *Cylactis* — CZAPIK (1981) regarded apomixis possible, though there was no direct evidence of successful apomictic reproduction.

After all, technical inconsistency may be regarded as the most likely explanation for the drupelet formation (cf. FOWLER and JANICK 1972). Since in these pilot studies, both the quality of the isolation and the purity of pollination in addition to the pollen and clonal sources were not always carefully controlled.

LARSSON (1969) also reports an exceptional case of a small amount of self-fertility towards the end of July in one year. She tried to check this in later years, using only very carefully isolated flowers, but not fruits ever developed.

At the Komarov Institute in Leningrad there is reputed to be a self-fertile strain of the arctic bramble (HÄRDH 1976), but I have not succeeded in obtaining any confirmation of this. Thus, until now there seems to have been no reproducible or confirmed cases of self-fertility in diploid *R. arcticus*.

In pollination experiments using parents and their  $F_1$ -progeny (TAMMISOLA and RYNNÄ-

NEN 1970), it was shown that the self-sterility is due to an oppositional type of an incompatibility system, controlled by a single gene locus and many alleles. The system was further shown to obey a sporophytic-gametophytic control of pollen tube growth. This means that the arctic bramble has a self-incompatibility system of the *Oenothera*-type, which is fairly widely distributed in the family Rosaceae. Typical of this system, the self-incompatibility (at least partially) no longer held true at the tetraploid level (TAMMISOLA and RYYNÄNEN 1970), though LARSSON (1969) reported the autotetraploids of *R. arcticus* to be self-sterile.

### 1.2. Variation of fruit set in nature

SAASTAMOINEN (1930) divided Finland into zones on the basis of the occurrence and productivity of the arctic bramble (Fig. 5). Across Central Finland, broadening towards the North in the Eastern coastal regions of the Gulf of Bothnia, should be its *optimal zone*. Though the arctic bramble could be found also to the South and to the North of that zone, in these areas it set fruit only occasionally and in small amounts, despite locally rich flowering.

Saastamoinen proposed several reasons for the recorded variation in occurrence and fruit set. The arctic bramble may be *too continental* for thriving in certain areas. In her opinion, *general temperature* maps seemed to correlate well with the zones introduced by her. M. J. Kotilainen (SAASTAMOINEN 1930) suggested that *too hot summers* make the flowers to dry. *Thick snow cover* might be profitable: it prevents an early start of growth, thus protecting the flowers against frost damage in spring — the best flowering period of the arctic bramble. In Southern Finland, *intensive farming* may have harmed it, by robbing it of suitable habitats; in addition, *heavy clay soils* are often encountered in the South — a soil type apparently poorly suited for *R. arcticus*.

The overall productivity of arctic bramble in Finland is well known to vary very much

from one year to the next. In essence, this variation must be due to different weather conditions. In addition SAASTAMOINEN (1930) noticed that there were also great differences in fruit set inside the zones. These *local differences* were attributed by her to differences in *moisture* and *illumination*. For good fruit set, adequate moisture is required. According to her, light, however, should not be a limiting factor, since even in shaded *Alnus* thickets, there is plenty of light in spring and early summer. ERVI et al. (1955) considered that arctic bramble might even favour slight shadowing.

ERVI et al. (1955) suggested that outside of the optimal zone, there may be *more severe competition* from the rich natural flora, as well as *scarcity of habitats* favourable to the arctic bramble. They also considered the possibility that the zonation might be based on the presence of *mycorrhizal fungi*. This hypothesis, however, was rejected on the basis of the results in their study. They further point to the possible role of *micro nutrients*. Using semi-quantitative analysis, no differences, however, could be recorded in the trace elements from fruiting versus non-fruiting shoots. Comparing the nutrients in the soil, they found however clearly more phosphorus, copper, zinc and manganese in heavily productive areas than in those areas which gave low yields.

Further explanations based on microclimate have been given. KOTILAINEN (1949) supposed that in spring and early summer, less *frost damage* occurs near the lake districts (cf. SOLANTIE 1976), which results in greater arctic bramble yields in these areas. Some evidence is given by ERVI et al. (1955) that in a plastic greenhouse, more fruits are produced. In a plastic greenhouse, HIIRSALMI (1971) obtained best yields in a mist compartment. He thus concluded that the weak berry production of the arctic bramble in Southern and South-Western Finland could be due to low *relative humidity* of the air in spring (which he believed to prevent the dehiscence of the anthers). This generalization to encompass



natural conditions may, however, be too far-reaching, taking into consideration the 'exotic' conditions prevailing in the greenhouse. For instance, the temperature was not controlled and accordingly it was much higher in the plastic greenhouse than in nature.

ERVI et al. (1955) note that the distribution of the *pollinating insects* could be of great importance concerning fertilization. As was already stated above, only honey bees and bumble bees are worthy of consideration.

The *honey bee* only occurs in Finland in domestic hives. About 80 per cent of the bee keepers in Finland are situated in five Southern or Southwestern provinces i.e. Kymi, Uusimaa, Turku & Pori, Häme and Vaasa (ANON. 1976, VARIS 1981). Bee keeping is rarely practised in other parts of the country. The approximate northern limit for bee keeping is from Kainuu (say, Suomussalmi, 65° N) in the East to Rovaniemi (66.5° N) in the West. The effect of a bee hive is, however, only local, since the normal flying range of the honey bee may be less than 1 km (RIBBANDS 1951, BEUTLER 1954), though even distances as much as 10 km have been reported (FRISCH 1959). The overall distribution of honey bee colonies in Finland is quite low, only about 0.09 colonies per square kilometre (VARIS 1981).

The densities of *bumble bees* are difficult to measure and vary greatly both annually and during the season, as well as according to locality, habitat or species. Hence, no overall estimates for bumble bees in different parts of Finland are available (PEKKARINEN et al. 1981, TERÄS 1983). However, on the basis of distribution information (PEKKARINEN and TERÄS 1977, PEKKARINEN et al. 1981) and sporadic studies (TERÄS 1983, 1985a), a rough general overview can be formed. Hence, bumble bees seem to be commonly found all over Finland. In N Finland, the species are very different from those in S Finland. Yet there is no indication that there should be a difference in the overall density of bumble bees between the North and the South. Only in

Lapland the densities of the relevant bumble bee species are known to be essentially lower.

The local abundance of bumble bee colonies is likely in part to be dependent on the availability of favourable nectar and pollen sources early in spring. Modern, intensive farming practices have greatly decreased the occurrence of many such plant species, e.g. willows, around and in the fields (PEKKARINEN et al. 1981, TERÄS 1985a, b). Bumble bees, however, are known to search for nectar and pollen at distances of more than a kilometre (TERÄS 1983).

Though comparable records are missing, there seems to be no reason to suppose that the northern bumble bee species should display an entirely different type of flower preference. Thus, supported by scattered records from nature, it can be concluded that the northern bumble bees keep visiting *R. arcticus* as well. Hence, variation in densities of pollinating insects can not generally explain differences in the fruit set, except perhaps in Lapland.

ZELLER (1964) studied the ontogenetic details in the overwintering buds of the arctic bramble. In her material, strains from the 'optimal' zone had a smaller *number of flower primordia* (1—3) than either the strains from South or from North Finland (3—6).

Another explanation also based on *strain differences* in different localities, had been proposed by M.Sc. (Agr.) M. Salminen (ERVI et al. 1955, VAARAMA 1965). He believed that self-fertile strains of the arctic bramble would occur frequently in Central and occasionally in Northern Finland, while they would seldomly be encountered in South Finland. Hence, the zonal — and also local — differences in fruit set reported by SAASTAMOINEN (1930) could be attributed to the distribution of self-fertile strains. In Southern Finland, due to the self-sterility of the prevailing strains, almost no berries would be produced. Hence, ERVI et al. (1955) propose that in further studies it should be made clear whether the strains in the 'optimal' zone differ from those

outside. TAMMISOLA and RYYNÄNEN (1970), using material consisting of only 16 strains, reported evidence contradictory to Salminen's hypothesis which has been presented above. Only a few strains showed signs of possible disturbances in self-sterility, and contrary to Salminen's hypothesis, these were from the 'poorly producing' zones instead of from the more richly producing ones. Thus, from the 'optimal' zone, only self-sterile strains were found.

SALMINEN (1948) and later on also LARSSON (1968) and Y. L. A. Mäkinen (see TAMMISOLA and RYYNÄNEN 1970) have suggested that certain cases of non-fruiting in natural occurrences of arctic bramble might be explained by unclonality and self-sterility. To formulate this hypothesis more generally: the *clonal composition of populations* is proposed as an explanation to the 'odd' variation in fruit set in natural *R. arcticus* populations in Finland.

### 1.3. Hypotheses and study plan

In order to ascertain the relevance of these last mentioned propositions, the present study was started. To begin with, the following *hypotheses* were formulated.

1) All or practically all clones are postulated to be *self-sterile*. If any (exceptional) self-fertile clones might occur, they are predicted to be distributed in a non-systematic way in relation to the productivity zones.

2) The local differences in fruit set are primarily due to a different *availability of successful pollen*.

If there is a sufficient number of foraging bumble bees or honey bees, the availability of successful pollen will primarily depend on the *number and spatial pattern* of the 'equivalence classes of incompatibility' in the population (TAMMISOLA and RYYNÄNEN 1970, TAMMISOLA 1981). Hence, vigorous, richly flowering

populations with *no fruit set* should generally be regarded as containing a single equivalence class. On the contrary, *richly fruiting* populations should as a rule display a mixture of many incompatibility classes. In theory, two classes would suffice, provided they were even in numbers and intimately intermixed. Vigorous populations with a *poor fruit set* should usually consist of only two or few equivalence classes. These classes should either be very uneven in numbers or be well separated spatially.

*In Lapland*, and perhaps in intensively cultivated areas outside the bee keeping area, there may occasionally also occur *local lack of effective pollinating insects*. One should, however, notice that in North Finland, early summer frosts are more common than in Southern Finland. Hence, loss of fruit crop may also be due to *climatic factors* (VAARAMA 1965).

To check these hypotheses, the following general *plan of study* was initiated. Samples should be taken from arctic bramble populations in different parts of Finland. These samples should be analyzed to yield the *number and spatial pattern* of the *equivalence classes* of incompatibility in each population. *Self-fertile clones* should be looked for in a large number of populations. This data should be compared to the *fruit set* information from the populations in question.

In this study, generally only vital populations — ie. vigorous both in growth and in flowering — were considered. In each of arctic bramble productivity zones (Fig. 5, SAASTAMOINEN 1930), fruiting as well as non-fruiting populations were included in the study. Thus it would be possible to reveal any *zonal differences* in the equivalence class structure of eg. non-fruiting populations, as well as in the occurrence of (speculative) self-fertile clones.

## 2. Material and methods

### 2.1. Sampling of populations

During the first four months of 1976, a questionnaire was distributed via mass communication media in Finland, asking for information about arctic bramble. Inquiries were made regarding in particular vegetatively vigorous and richly flowering but yet fruitless populations. During 1976—1977, more than 300 persons responded to the questionnaire. On the basis of this accumulated information, the populations for the present study were chosen. Adjusting the study plan (see 1.3.), in each of the four “productivity zones” of arctic bramble (Fig. 5, SAASTAMOINEN 1930), both non-fruiting, poorly fruiting and richly fruiting populations were to be investigated. Within each zone and in each fruiting class, these populations were acquired through random sampling.

During the summer 1976, collection tours to some of these populations were made. In spite of careful checking of the information in advance with the correspondents, a few of the populations first occurring in the collection scheme proved *in situ* to differ too much from the reported fruiting class, thus rendering modification of the collection scheme necessary. For practical reasons, the populations which were visited, were as a rule included in the collection scheme, but now represented their actual fruiting class. In 1977, a revised collection scheme was made: the still missing populations of each fruiting class were selected at random. Also in 1977, a couple of originally selected populations had to be replaced by other ones, due to their wrong fruiting class or insubstantial growth area. The smallest growth areas were already excluded prior to random sampling — the minimum acceptable

size of the longest diameter was set to approximately 10 metres.

A total of 29 populations were taken (Fig. 7, Tab. 1).

### 2.2. Sampling of shoots

From each population, several shoots were sampled at random. The rootstocks of these shoots were dug up and sent for planting at our experimental field in Viikki, Helsinki. On the basis of this ramet sample (for the terminology, see TAMMISOLA 1986), the equivalence class pattern of the population could be analyzed.

#### 2.2.1. Number of ramets to be taken in a population

In order to study as many as 29 populations, the number of ramets from each population cannot be too exhaustive. This number should, however, be large enough to reveal at least “essential” differences between populations, with regard to the number and spatial pattern of their contained equivalence classes of incompatibility (see 3.1.). *Thirty ramets* were considered to be enough, on the basis of following reasons.

1. Postulating *m* equally frequent equivalence classes in (an infinitely) large population of shoots (cf. TAMMISOLA 1986, ‘Richness’), the problem concerning the number of shoots to be taken can be studied via the theory of ‘occupancy’ (see RIORDAN 1958, p. 90).

Taking shoots from the population (ie. from an array of *m* equivalence classes) is analogous to putting objects into *m* cells, each with equal probability. Combining his results (RIORDAN 1958, p. 90—91), probability of no

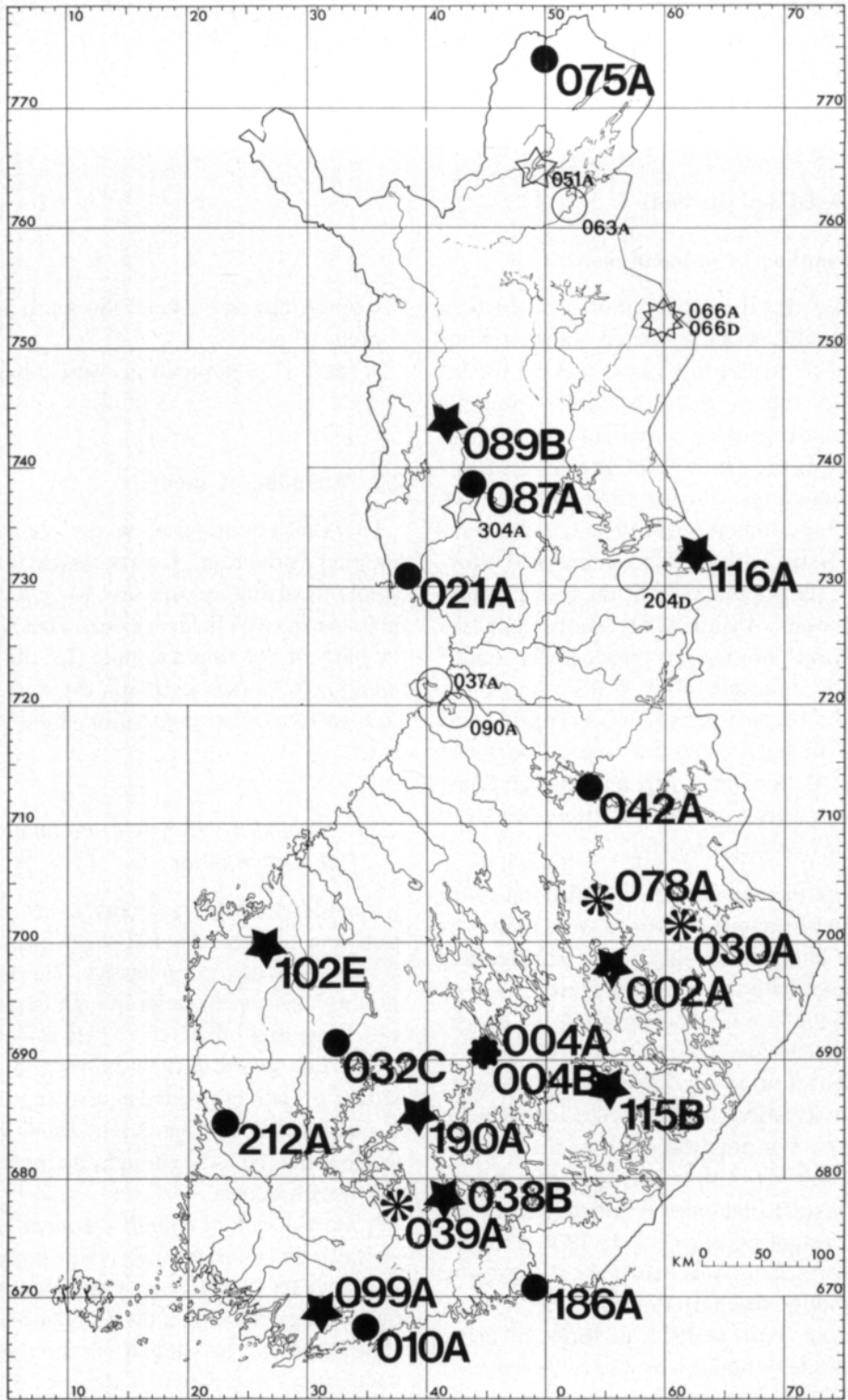


Fig. 7. Arctic bramble populations chosen for the present study. The populations which remained (almost) totally unanalyzed into equivalence classes, have been denoted by open symbols and small type. Fruit set is indicated: dot = rich, asterisk = poor, star with five points = none.

Table 1. Arctic bramble populations chosen

Popul. label <sup>1</sup>	Fruit set <sup>2</sup>	Prod. zone <sup>3</sup>	Sampled area [m × m]	Popul. continues <sup>4</sup>	Longitude ° ' E	Latitude ° ' N
002A	0	I	8 × 14	no	28 06	62 54
004A	2	I	1 × 25	int.	25 58	62 16
004B	1 <sup>a</sup>	I	4 × 26	no	25 58	62 16
010A	2 <sup>a</sup>	III	9 × 14	yes	24 15	60 09
021A	2	I	2 × 21	yes	24 29	65 52
030A	1	I	8 × 9	int.	29 15	63 13
032C	2	II	9 × 18	no	23 38	62 17
(037A)	2	I	12 × 14	yes	25 02	65 02
038B	0	IV	13 × 14	no	25 26	61 10
039A	1	III	4 × 21	no	24 40	61 04
042A	2	II	7 × 10	yes	27 45	64 16
(051A)	0	III	8 × 32	int.	26 44	68 55
(063A)	2	III	9 × 13	yes	27 32	68 40
[066A]	0	III	8 × 18	yes	29 22	67 48
[066D]	0	III	5 × 25	yes	29 22	67 48
075A	2	IV	3 × 25	no	26 57	69 44
078A	1	I	8 × 13	no	27 49	63 26
087A	2	II	10 × 11	int.	25 38	66 35
089B	0	II	4 × 18	yes	25 07	67 03
(090A)	2	I	5 × 23	int.	25 24	64 51
099A	0 <sup>a</sup>	III	6 × 21	no	23 35	60 18
102E	0	I	5 × 26	no	22 19	62 59
115B	0	I	11 × 14	no	28 01	61 59
116A	0	III	10 × 14	yes	29 50	66 01
186A	2	III	9 × 10	int.	26 52	60 29
190A	0	IV	5 × 21	int.	24 55	61 49
(204D)	2	III	2 × 23	int.	28 44	65 55
212A	2	III	8 × 13	yes	21 54	61 38
[304A]	0	II	15 × 20	int.	25 29	66 24

<sup>1</sup> Due to their weak condition in the experimental field, populations enclosed in square brackets [ ] remained totally (and those in parentheses ( ), almost totally) unanalyzed into equivalence classes.

<sup>2</sup> '0' = none, '1' = poor, '2' = rich; representing less than 1, 1–15, and more than 15 'full' berries (ie. containing at least 10 drupelets) per a totally covered (100%) m<sup>2</sup> of arctic bramble vegetation.

<sup>3</sup> See Fig. 5.

<sup>4</sup> 'yes' = arctic bramble occurs almost continuously in at least about two times of the area sampled; 'int.' = population continues but intermittent, ie. with at least about a 20 m break.

<sup>a</sup> Date of collection was too early for reliable *in situ* records of fruit set. Hence, classification is primarily due to the previous and posterior reports of the correspondents.

empty cells remaining after putting n objects at random into m cells will be

$$(1) P(n, m; m) = \frac{m!}{m^n} \cdot S(n, m),$$

where S(n, m) is a Stirling number of the second kind. This formula will therefore also tell the probability of getting all m equivalence classes represented in a sample of n shoots from the population.

These Stirling numbers have been tabulated, eg. in RIORDAN (1958) up to 10 and in DAVID et al. (1966) up to 25. Unfortunately, there were no tables that could have been used for a sample size greater than 25. Therefore, an expression for the calculation of the desired probability was derived

$$(2) P(n, m; m) =$$

$$1 - \sum_{i=1}^{m-1} (-1)^{i-1} \cdot \binom{m}{i} \cdot \left(1 - \frac{i}{m}\right)^n.$$

Probabilities up to the sample size of 60, which were calculated utilizing this expression, are presented in Table 2.

By using a sample size of 30, it is almost certain to obtain all of the equivalence classes present in the population, whenever the population contains at most 6 different equivalence classes. The risk of losing any

of the equivalence classes will then be less than 3 per cent. Using a *sample size of 15*, the risk of losing equivalence classes were considerably greater: 36 per cent for populations with 6 classes. At a risk of 5 per cent, all classes can be guaranteed in a sample only in populations with at most 4 classes. At this risk level, a *sample size of 60* would suffice for populations containing at most 11 classes.

Since, however, the amount of labour required in the analyses will increase very rapidly with the number of ramets to be studied (see 3.), a sample size of 60 is far too large to be used in the study. A sample size of 30 is still manageable, though only via the careful optimization of procedures as applied in this study.

It can be shown that  $P(n,m;s)$  (ie. the probability of getting exactly  $s$  equivalence classes, out of a total  $m$  in the population, into a sample of size  $n$ ) can be traced back to the expression (2), viz.

$$(3) \quad P(n,m;s) = \binom{m}{m-s} \cdot \left(\frac{s}{m}\right)^n \cdot P(n,s;s)$$

With reference to a population containing  $m=11$  equivalence classes, the probability distribution of the number of equivalence classes occurring in a sample of size  $n=30$ , is given below:

N:o of classes	1-6 pooled	7	8	9	10	11
Probability	.000006	.0004	.01	.10	.40	.49

This example shows that almost always, nearly all of the classes will be contained in the sample — just as could be supposed on the basis of the high (10.4) number of expected classes (Table 2).

According to RIORDAN (1958, p. 101), the expectation of the number of equivalence classes,  $s$ , occurring in a sample will be

$$(4) \quad E_{n,m}\{s\} = m \cdot \left[1 - \left(1 - \frac{i}{m}\right)^n\right].$$

These expected numbers of revealed classes have also been presented in Table 2.

II. Postulating  $m$  *unequally frequent equivalence classes* in (an infinitely) large population of shoots, the problem of sample size turns out to be much more difficult. It is, however, still simple to study the most severe type of under-representation in a sample,

Table 2. Probability of acquiring, into a sample containing  $n$  ramets, every one of the  $m$  equivalence classes present in the population, provided all the classes are equally frequent.

	Sample size (n)									
	5		10		15		30		60	
	P <sup>a</sup>	E <sup>b</sup>	P <sup>a</sup>	E <sup>b</sup>	P <sup>a</sup>	E <sup>b</sup>	P <sup>a</sup>	E <sup>b</sup>	P <sup>a</sup>	E <sup>b</sup>
m =										
2	.94	1.9	.998	2.0	1.00	2.0	1.00	2.0	1.00	2.0
3	.62	2.6	.95	2.9	.99	3.0	1.00	3.0	1.00	3.0
4	.23	3.1	.78	3.8	.95	3.9	.999	4.0	1.00	4.0
5	.038	3.4	.52	4.5	.83	4.8	.99	5.0	1.00	5.0
6			.27	5.0	.64	5.6	.97	6.0	1.00	6.0
7			.10	5.5	.43	6.3	.93	6.9	.999	7.0
8			.028	5.9	.25	6.9	.86	7.9	.997	8.0
9			.0047	6.2	.12	7.5	.76	8.7	.99	9.0
10			.00036	6.5	.046	7.9	.63	9.6	.98	10.0
11					.014	8.4	.49	10.4	.96	11.0
12					.003	8.7	.36	11.1	.94	11.9
13					.0006	9.1	.24	11.8	.90	12.9
14					.00006	9.4	.15	12.5	.84	13.8
15					.000003	9.7	.088	13.1	.78	14.8
16							.046	13.7	.70	15.7
17							.022	14.2	.62	16.6
18							.010	14.8	.53	17.4
19							.004	15.2	.45	18.3
20							.001	15.7	.36	19.1

<sup>a</sup> Probability  $P(n,m;m)$  (see expression 2).

<sup>b</sup> Expectation of the number of equivalence classes in the sample, that is  $E_{n,m}\{s\}$  (see expression 4).

which would lead to a critical misinterpretation of the equivalence class structure of the population. That is, if the sample of shoots happens to contain only one equivalence class, in spite of several which occur in the population, the population will be fatally *misclassified*, regarding the central problem of the study.

Denoting the relative frequencies of the equivalence classes in the population by  $p$ , the *probability to misclassify* the population as uniclassical will be

$$(5) P_{\text{mis}} = p_1^n + p_2^n + \dots + p_m^n \leq \epsilon,$$

where  $\sum p_i = 1$  and the 'greatest acceptable' value of this probability of misclassification has been denoted by  $\epsilon$ . This expression is known to yield its minimal value when all equivalence classes are equally frequent in the population. Whenever one of the classes greatly predominates in the population, while the frequencies of the other classes approach a value of zero, the expression will on the contrary tend to its maximal value of 1. Even in this case, however, by increasing the sample size  $n$ , the probability of misclassifying is diminished to an acceptable level.

In an uneven population where one equivalence class predominates, the probability of misclassifying the population as uniclassical will be worst if  $m = 2$ , ie. if the population contains only *two equivalence classes* of incompatibility. Misclassification probabilities for such populations ( $m = 2$ ) are presented in Table 3. The *probability of misclassification*,  $P_{\text{mis}}$ , can be brought *below .05* if and only if sample size used is at least as high as that given below:

Table 3. Probability of acquiring, into a sample containing  $n$  ramets, only (either) one of the two equivalence classes constituting the population.

	Frequency of the more common equivalence class						
	.5	.6	.7	.8	.9	.95	.99
$n =$							
2	.500	.520	.580	.680	.820	.905	.980
4	.125	.155	.248	.411	.656	.815	.961
6	.031	.051	.118	.262	.531	.735	.941
8	.0080	.018	.058	.168	.430	.663	.923
10	.0020	.0062	.028	.107	.349	.598	.904
15	.00006	.0005	.0047	.035	.206	.463	.860
30	$2 \times 10^{-9}$	$2 \times 10^{-7}$	.00002	.0012	.042	.215	.740
60	$2 \times 10^{-18}$	$5 \times 10^{-14}$	$5 \times 10^{-10}$	$2 \times 10^{-6}$	.0018	.046	.547
299							.0495

Frequency of commonest class	.5	.6	.7	.8	.9	.95	.99
Sample size required	6	7	9	14	29	59	299

Hence, considering fatal misclassifications, the proposed sample size of 30 would suffice well for populations so uneven with regard to the frequencies of shoots in their equivalence classes (see TAMMISOLA 1986 for 'evenness') that their most frequent equivalence class will contain up to 90 per cent of the total number of arctic bramble shoots in the population. In a population with a still more prevalent equivalence class, many more ramets would be needed in a sample to attain at a considerably sure classification of the population.

Regarding the problem of fruit set in the population, such an extremely 'uneven' population will only slightly differ from a purely uniclassical population, since the average availability of successful, 'alien' pollen in such a population would be negligible. Hence, in practice, such a misclassification of an extremely 'uneven' population would scarcely interfere with the results of the study.

As already stated above, in populations with more than two equivalence classes, a somewhat smaller sample size than above will suffice. Omitting the postulate of infinitely many shoots in the population will also diminish the sample sizes required.

### 2.2.2. Randomization

Since the problem of fruit set concerns, in the short term, only the shoots in a *generative state* (see TAMMISOLA 1986), then only such shoots were sampled.

In order to provide *each shoot with equal probability*, the generative shoots would be numbered in sequence, and then — based on these numbers — the shoots could be chosen randomly using a table of random numbers. In practice this is impossible for arctic bramble, due to the very great number of shoots — too much labeling work would be needed, which in addition would result in the trampling of the plants.

Alternatively it can be decided to apply the probability *equally* to the *area* inhabited by the population. Using this method, individual shoots from dense patches would be less likely chosen than shoots from less crowded areas. This coverage effect could be avoided by defining that a shoot is recorded as being selected only if the random co-ordinates fall exactly on the place of *stem base* on soil surface. This spot is, however, very small, which would result in an untenable frequency of misses — the great majority of chosen random co-ordinates would not hit any shoot at all, resulting in immoderate trampling.

To remedy this defect, the co-ordinate may be increased to encompass a small area — then the hits will be common enough. Hence, a shoot is considered being selected, if the random co-ordinates fall within a *circle of a fixed radius* surrounding the point of the stem base on the soil surface. When several shoots are selected simultaneously, then the one situated nearest to the random co-ordinate, would be chosen. Due to the fixed circle area, shoots from the *less dense parts* of the population would be *over-represented* in the sample. Therefore, the shoots in the sample might not unbiasedly represent the quantities of different pollen classes in the pollen pool of the population. Nevertheless, this method would likely yield a *better representation* than shoot number based methods, of the *number of equivalence classes* present in the population.

Actually, this last method of randomization was applied in sampling from the natural populations in the present study. On the basis of pilot experiments in nature, the radius was fixed to be 50 cm.

A straightforward technique to transfer random co-ordinates onto the population area, would be ordinary triangular measurement. With a fixed base line, the point can be transferred into the field, by the use of two measuring bands. This technique would require two persons — and if employed by a single person, it would involve too much trial and error and consequent trampling of the sampling area. Shrubs etc. make the situation still worse. Hence, this technique was generally not applied in the randomization step. In a later stage, when *mapping* the sampled shoots, such triangle measurement was however used.

Suggested by pilot experiments, a technique applying polar co-ordinates and a theodolite, was constructed. Polar co-ordinates could easily be transferred into the field by a single person in a single trial. This was made possible by modifying the theodolite in such a way that it could be used in a reverse direction as well — a stick was attached along its telescope, with nails erected at both ends. This technique resulted on average not more than about 10 cm error, over the distances of about 20 metres. Even this error does not matter, since it is randomly distributed.

Random numbers were drawn from tables of FISHER and YATES (1957), providing 15 000 numbers, as well as from those of ARKIN and COLTON (1963), providing 8 000 numbers.

### 2.3. Observations in natural populations

As already described, 30 random shoots were acquired in each population, and were planted at Viikki for further investigations. Their exact co-ordinates in the population were measured. Before digging them up, the surrounding 1 m<sup>2</sup> of each shoot was studied. The cover percentage of the above ground parts of arctic bramble was subjectively estimated visually, as usual. Then the number of 'big' berries, consisting of at least 10 drupelets, as well as that of 'small' ones, consisting of from 1 to 9 drupelets, were counted. Furthermore, the number of buds, flowers, and degenerating old flowers (ie. not yielding any drupelets) were counted.

These observations were made at different times at different locations. For practical reasons, they had to be made during the collection trips, which occurred during June, July and August in 1976, and during June and July in 1977. Hence, the data in different populations will only be roughly comparable. The recorded numbers will, however, serve as a much more reliable basis for the classification of populations with regard to fruit set, than did the reports of the correspondents. Unfortunately, a couple of populations (eg. 010A) had to be visited at the beginning of the flowering period — too early for the fruit set data to be trustworthily recorded. In such populations, classification with regard to fruiting had to be made almost solely on the basis of the previous and later reports of the correspondents.

### 2.4. Management of the sampled ramets

After recording the observations in a population, the sampled shoots were dug up, in a clod of soil of about 30–40 cm in diameter and 30 cm deep. Immediately after digging, the bulk of the soil was carefully removed from the clod. The clod was moistened and packed with its contained shoots into a plastic bag. The bags were then transported to Viikki. In the worst cases, the transport lasted for several days, resulting in the shoots arriving at Viikki in a poor condition. After arrival, the clods were entirely broken down, and the rhizome containing the actually sampled shoot was planted in a pot or in the field.

Care was taken to acquire only a *single* rhi-



zome. If the shoot and rhizome in question were in a considerably weak condition, a vice-rhizome from the same clod was planted in another pot. In 1976, all sampled shoots were planted in pots in the greenhouse. Those collected in 1977 were planted directly into the experimental field (see below); the vice-rhizomes of them were however managed in pots in the greenhouse.

## 2.5. Greenhouse in 1976—78

In the greenhouse, arctic bramble plants were cultivated in pots on growth peat (Fig. 2). In the autumn and winter, extra light was provided daily for several hours by high pressure halogen lamps. The plants had a normal winter rest of about three months.

In 1976, a ready fertilized peat was used. Unfortunately the fertilizer levels varied considerably, cross unbalance or deficiency of nutrients causing severe stress to most of the transplants. Especially during the winter, many of them died — a loss of up to 30 per cent was recorded. One other reason for this death rate was that the greenhouse could not be kept cold enough to provide a suitable environment for dormancy. During the winter time, a couple of voles also succeeded in entering the greenhouse, and caused much damage by eating the resting buds in many of the pots.

Also in the autumn and winter, before settling into dormancy, two rust fungus species, *Pucciniastrum arcticum* (Lagerh.) Tranzsch. and *Phragmidium arcticum* Lagerh. (see MÄKELÄ 1984), caused damage to the weakest specimens. While rarely encountered in natural populations (075A and 102E), the third rust fungus, *Gymnoconia peckiana* (Howe) Trott. was never recorded in the greenhouse, nor in the experimental field. This was fortunate, since it seriously deforms the attacked shoots, rendering flowering or at least fruit set impossible. In winter, a powdery mildew fungus (*Sphaerotheca aphanis* (Wallr.) Braun) was recorded on a few individuals (the species was kindly determined by M.Sc. (Hort.) P. Alanko). Later on it was recorded occasionally in the field, especially on the flower stalks and flowers of certain clones, and diminished their fertility. This fungus has only been found on arctic bramble once before (VARDJA and KARIS 1985).

From the summer 1977 onwards, all fertilization of peat was undertaken by the Department of Plant Breeding. The rhizome was planted in a peat fertilized with 6 kg of dolomite chalk, 200 g of triple phosphate and 800 g of Peat Y fertilizer (containing 11 % N, 11 % P, 18 % K, 0.1 % B, 1.1 % Cu, 0.5 % Mn, 0.95 % Fe, 0.5 % Zn and 0.1 % Mo) per one m<sup>3</sup> of ready peat. Some weeks later, the transplants were transferred into a similar peat which contained a double dose (1600 g) of Peat Y fertilizer. From 1977 onwards, also the overwintering temperature could be lowered near to the freezing point,

providing a more satisfactory winter dormancy period. Accordingly, a much better result than in 1976 was achieved, with losses of about a few per cent.

## 2.6. Experimental field in 1977—80

In the spring of 1977, an experimental field for arctic bramble was established at Viikki. It was a little extended in 1978, totalling in about 20 ares and 1225 plots (App. A). Plot size was 75 × 75 cm, and each plot was surrounded by a 50—100 cm zone of free space. Unfortunately, no under soil isolation between the plots could be financed. The 30 specimens of a population were planted into a block of 5 rows × 6 columns. In addition to the samples of the populations under study, the hundreds of single accessions sent by the correspondents in the arctic bramble survey, were also planted. In 1977, the effect of planting time was clearly visible. Shoots planted towards the end of July, suffered more than the ones planted in June. This 'planting effect' lasted over many years, and resulted in slower growth in the respective plots.

The area consisted of sandy soil, with a moderate humus content and pH ranging from 5.7 to 6.2. In the beginning, 1140 kg of Garden Y1 fertilizer (containing 10 % N, 4 % P, 17 % K, 2.5 % Mg, 0.15 % B, 0.4 % Cu, 0.7 % Mn, 0.1 % Fe, 0.03 % Zn and 0.02 % Mo) and 5400 kg of dolomite chalk was applied per hectare. Furthermore, the area was made more suitable for arctic bramble by adding 90 m<sup>3</sup> of peat to it. Each m<sup>3</sup> of the peat contained 1.6 kg of a fertilizer with 9 % N, 11 % P, 16 % K, 0.1 % B, 0.2 % Cu, 0.5 % Mn, 1.0 % Fe, 0.3 % Zn, 0.1 % Mo, 0.2 % Na, 3.9 % S and 3.5 % Ca; furthermore, 10 kg of dolomite chalk was added. In the plots, a 2—3 cm deep layer of sand (with grain size of less than 4 mm) was used as a mulch. In the first two years, weeds were controlled by spraying herbicides (paraquat and glyphosate) in the field. Since the stands in the plots were still small in area, they could be protected during spraying by placing buckets upside down over the plants. In later years, weeds in the plots were controlled by hand. Details of a cultivation technique have been described by RYNNÄNEN (1971, 1973).

In the field, clones slightly infected by either of the two rust fungi mentioned above (*Pucciniastrum* and *Phragmidium*), were occasionally found. In late summer, powdery mildew (*Sphaerotheca*) which originated from the greenhouse, was recorded on a couple of populations. Fruit set was, however, not impaired, since the bulk of the flowers had already been produced before the end of June and hence avoided the infection. In plots with weakly growing or withering shoots, several pathogenic or slightly pathogenic fungi were found (at the base of the stems and

in the roots or on the canes), including *Cylindrocarpon destructans* (Zins.) Scholten, *Fusarium* and *Phoma* species and a *Didymella* species (RUOKOLA 1981). No fungicides were applied, since the fungus attack appeared to be moderate enough not to threaten the present population study.

In 1978, the stands in the plots were already large enough for a medium scale controlled crossing experiment, and for some pilot studies concerning the recursive system of experiments and analyses (see 3.). Preliminary observations on vitality, phenology, morphology and fruit set were also made.

In 1979, the bulk of the crosses were made. This large scale crossing work was now, even in details, guided using a computer system (see 3.3.), the construction of which was completed during the winter of 1978—79. In 1979, numerous observations were made about the characteristics important in fruit set and cultivation.

In 1980, crossing was still continued on a limited scale. It was confined to such areas in the experimental field, where different stands had apparently not yet intermixed. Strict care was taken to utilize only the flowers in the centre of each plot. In addition, the morphological characteristics of the shoots were continuously monitored, in order to reveal any cases of a hidden intrusion.

## 2.7. Techniques in crossing

In order to minimize the possibilities of mistakes, and also to cope with the large scale of the crossing program in the present study, the earlier methods used in crossing (TAMMISOLA and RYYNÄNEN 1970) were now refined.

Buds of *R. arcticus* were isolated into bags well before the opening of the bud. The best time for that was, when the buds are about 2—3 mm, i.e. some days before their opening. In a few exceptional clones, especially those infected by the powdery mildew, buds had to be isolated when they were much smaller, since in them the buds were partially opened (a 'hole' appeared) at an early stage.

Isolation bags were specially constructed for the study: one side was made out of a woven cotton fabric, while

the other side was made out of a plastic film. The two sides were glued together on their edges by a hot pressure. By utilizing such bags, it was possible to visually inspect the buds to see when the right stage had been reached for performing the crosses. Hence, a remarkable saving in time and effort was achieved.

In practice, the optimal stage for a flower bud to be used as a *female* part in a cross, will be one day prior to its anthesis. Morphologically, the petals are then already well visible, being about 2/3 of their final size; the bud has not yet opened, but it is still more or less cone-shaped. At this stage the stigmas will already be quite receptive, but usually not a single anther has opened yet. In spite of the self-incompatibility, the flowers to be used as female ones, were always carefully *emasculated*. Due to the flower morphology (Fig. 6), emasculation had to be done also to make an even pollination of all the pistils technically possible.

Emasculation technique remained unaltered. Flowers of arctic bramble, while being perigynous, can be emasculated via cutting by a razor blade around the bud, between its gynoecium and whorls of stamens. The 'cap' containing the perianth and the whorls of stamens, will then be removed, leaving the gynoecium naked.

Flowers to be used for *pollen*, were cut and placed in labelled petri dishes. The best time for this would usually be immediately at, or some hours after anthesis. The uppermost whorls of stamens would just have opened the pollen sacs, and the flower would contain much fresh pollen. If required, somewhat older flowers — up to two days after anthesis — can be utilized for pollen (then only fresh pollen from the innermost, latest opening whorls can be used).

Fresh pollen is pale yellow and sticky, while old pollen with questionable fertility is white and dusty in appearance. Pollen moistened with water or nectar will lose the fertilization capacity, and characteristically turns brown in colour. Threadlike structures among the pollen indicate a powdery mildew infection, which usually decreases the fertility of the pollen substantially. In the flowers to be used for pollen, petals were removed. Using fine forceps in a scissor-like action, a piece of the flower — with some sepals and a piece of stamen whorls but without any pistils — was cut apart. Taking this piece in the forceps, pollen is carefully but gently 'brushed' directly from the pollen sacs onto the stigmas of an emasculated flower. Provided the flower was emasculated at a right stage, pollination can be performed immediately after emasculation.

Depending on the amount and quality of pollen, usually 1/4 to 1/2 of the pollen flower was used for one pollination. The rest of the pollen flower was therefore used in a *control* cross and in one or two actual crosses. Controlling — via a cross — the quality of pollen in each utilized pollen flower separately, is most essential, since pollen quality varies greatly between flowers, even after visual checks have been performed.

About ten to fourteen days after the cross was made, the results were recorded. The computer system (see Fig. 8 below) provided in advance a list of the crosses to be inspected (App. C). In the pollinated flower, the number of ovaries beginning to enlarge was counted. For further technical details in crossing, see App. F.

## 2.8. Cytological staining method

The chromosomes of *R. arcticus* are small and quite difficult to stain. A Feulgen-Giemsa double staining method developed by Dr. S. Nokkala was chosen (see KOTIMÄKI and HIIRSALMI 1979). Even in the early stages of meiosis, this yields a good stain.

## 2.9. Statistical test methods

Most variables in the study were measured at a precision of a ratio scale. For some quantities, an ordinal scale or even a nominal scale was used. All statistical tests were therefore made utilizing *non-parametric* tests. Hence, there was no need for various (in principle dubious) transformations, which are commonly applied in order to achieve a normal distribution. In addition, a comparison of different variables became easier, when a common test with an equal power was used for each of them. Testing was performed by using the statistical program package SURVO84C, which has been developed in the Department of Statistics at the University of Helsinki. From the non-parametric tests available in the package, always the same and most powerful one was chosen whenever possible. In pairwise comparisons of means, the *Fisher-Pitman randomization test* was generally used (eg. CONOVER 1971, p. 328), and in exceptional cases *Mann-Whitney U-test* (CONOVER 1971, p. 216, SIEGEL 1956, p. 116). For a comparison of several means, *Kruskal-Wallis variance analysis* for ranks was chosen (CONOVER 1971, p. 229). When measuring an association (correlation) between two variables, *Kendall's* rank correlation coefficient ( $\tau$ ) was used (eg. SIEGEL 1956, p. 213).

### 3. Recursive system of experiments and analyses

To diminish the demand for human labour during the short active crossing period in summer, and to direct the choice of forthcoming crosses to the most relevant ones, a computer aided system was developed (Fig. 8). Knowledge about crosses made currently (App. B), and newly recorded results from previous crosses (App. C), were given as an input into the computer. As outputs, the system yielded a) lists of previous crosses, whose results were ready for recording (App. C), and b) present analyses of equivalence class structure of populations (App. D), and c) recommendations for new, most likely informative crosses to be made in next few days (App. E).

During the summertime, this whole procedure was repeated recursively. The system was run *one cycle per week*. Hence, results from all previous crosses could be utilized almost immediately (with a lag of 10–14 days from

the crossing date before the cross results could be recorded) for population structural analyses, and for directing the choice of new crosses to be made.

A more detailed description of the most essential steps in the recursive system, will be given later on, in the respective chapters.

Automated data processing was accomplished on a Burroughs B7800 computer at the Computing Centre of the University of Helsinki. Most computer programs on the system were self-tailored from the beginning. In the construction of the data collection and management system, valuable contributions were made by the Computing Centre, but the most important, analytic procedures were designed and coded by the author. The programs were almost exclusively coded using B7800 versions of FORTRAN IV and FORTRAN 77 languages.

The principal systems analysis phase took place in 1977–78, and the programming phase in 1978–79. The system was ready for testing in the summer of 1978, and was fully operational, i.e. with full guidance, from the summer 1979 onwards. The iterative ‘post-analysis’ system (see 3.3.6, 3.3.7) was designed and operational after completing the crossing phase, mainly in 1981–82.

#### 3.1. Gametophytic incompatibility yields equivalence classes

In a one locus *gametophytic* system of self-incompatibility, the incompatibility reaction will be determined by the incompatibility allele of the individual (haploid) pollen grain (see eg. ARASU 1968, TAMMISOLA and RYNNÄNEN 1970). The pollen tube growth of the

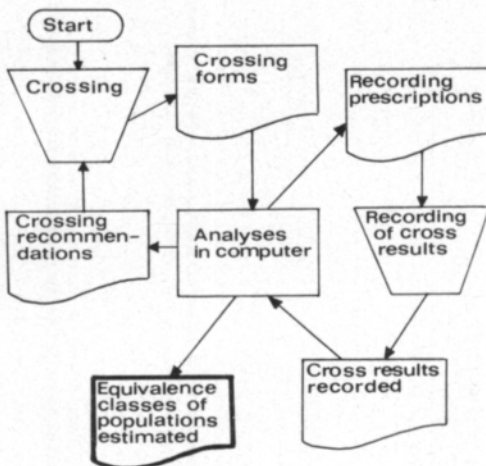


Fig. 8. Computer aided analysis and guidance system for equivalence classificatory crossing work. During the summer months in 1978–80, the system performed a cycle once a week.

grain will be inhibited in a (diploid) stylar tissue containing the same allele of the *s*-locus as the pollen. Since in arctic bramble, self-incompatibility is strict enough (TAMMISOLA and RYNNÄNEN 1970), and in addition, inbreeding depression is severe (see TAMMISOLA 1981), we can postulate that in practice, no *s*-allele homozygotes will occur in its populations.

Therefore, in crossing two arctic bramble ramets, we will as a rule have only the following three possibilities, the ramets having either 0, 1 or 2 *s*-alleles in common:

- a)  $s_1s_2 \times s_3s_4$  (ramets have no *s*-alleles in common), or
- b)  $s_1s_2 \times s_2s_3$  (ramets have one common *s*-allele), or
- c)  $s_1s_2 \times s_1s_2$  (ramets have two *s*-alleles in common).

In case a), all the pollen will be capable of fertilizing. In case b), only one half of the pollen will grow, i.e. only the grains containing  $s_1$ . However, since in artificial crosses, there is always much pollen on each stigma, fruit set is likely to be (almost) as good in case b) as in case a). On the other hand in case c), all the pollen grains will be inhibited, and accordingly no berries except for the occasional drupelets are to be expected.

Next, let us in a set of ramets define a relation  $R = \text{'the seed parent * is incompatible with the pollen parent *'}$ . This relation is reflexive, i.e.  $xRx$ , since arctic bramble is self-incompatible. It is also symmetric, i.e.  $xRy$  implies  $yRx$ , since *s*-allele homozygotes were *outruled*. A ramet homozygous at the *s* locus would ruin the symmetricalness:  $s_1s_2 R s_1s_1$ , but  $s_1s_1 R s_1s_2$ , since the pollen grains with genotype  $s_2$  would be able to grow uninhibited and hence to fertilize.

Our relation  $R$  defined above will be also transitive. This will hold even with *s*-allele homozygotes occurring.

Namely, regarding any two *s*-alleles (say  $s_1$  and  $s_2$ ), there could occur only five basic cases of  $R$ -related ramets, i.e.  $s_1s_1 R s_1s_1$ ,  $s_2s_2 R s_2s_2$ ,  $s_1s_2 R s_1s_2$ ,  $s_1s_2 R s_1s_1$ , or  $s_1s_2 R$

$s_2s_2$ . Looking through all the seven possible chain combinations that can be built out of these — that is

$$x = s_1s_1, y = s_1s_1, z = s_1s_1;$$

$$x = s_2s_2, y = s_2s_2, z = s_2s_2;$$

$$x = s_1s_2, y = s_1s_1, z = s_1s_1;$$

$$x = s_1s_2, y = s_2s_2, z = s_2s_2;$$

$$x = s_1s_2, y = s_1s_2, z = s_1s_2 \text{ OR } s_1s_1 \text{ OR } s_2s_2$$

— one can see that in each case,  $xRy$ ,  $yRz$  will actually imply  $xRz$ .

Hence, the relation  $R$  under study is an *equivalence relation* (see App. G). Therefore, in a plant species with a one locus gametophytic self-incompatibility system, *crosses* can be utilized in *subdividing* a population into equivalence classes. Shoots (ramets) that constitute an equivalence class, will not set fruit when pollinated with each other in any pairwise combination. On the other hand, any two ramets taken from different equivalence classes, will fertilize each other. Hence, these '*equivalence classes of incompatibility*' will in the most relevant way represent the situation in the population, with regard to possible sources of successful pollen.

In an oppositional one locus *sporophytic* system of self-incompatibility, the reaction of a (haploid) pollen grain is determined by the *s*-genotype of the (diploid) pollen parent, not by the *s*-genotype of the pollen grain itself. Growth and even germination of pollen is inhibited in all pistils which have an active incompatibility allele in common with the pollen parent. Dominance can occur, and the dominance reactions between *s*-alleles may even be different in stamens and in pistils (ARASU 1968). Hence, *asymmetric* reactions can occur (cf. Fig. 13). Therefore, in a sporophytic self-incompatibility system, the relation  $R$  defined above ( $R = \text{'the seed parent * is incompatible with the pollen parent *'}$ ), will generally not be an equivalence relation, and will therefore *not* provide us with equivalence classes.

*Self-compatible* ramets, being not even reflexive, would ruin the classification. If such ramets occurred, they should be left out of the equivalence class analysis. Regarding fruit set in the population, such ramets might, however, have a great influence.

*Homozygous s*-genotypes could in practice — in an order of decreasing likelihood — a) be misclassified into one of the equivalence classes, or b) be misclassified as a new, separate equivalence class, or c) erroneously combine together different equivalence classes which contain its *s*-allele, or d) erroneously

subdivide existing equivalence classes which contain its s-allele. Thus, counteractions against such severe errors ought to be made, eg. looking carefully for any morphological signs of inbreeding depression, and for inconsistent results in crosses. Any suspected cases should be inspected using reciprocal crosses — a difference between their results would reveal a s-homozygous ramet.

### 3.2. Deterministic model of incompatibility

In the bulk of the literature concerning incompatibility, cross results have been expressed only *qualitatively*, ejecting one symbol for a 'positive result' and another for a 'negative' result (eg. Fig. 13). Such commonly occurring qualitative tables tend to give the wrong impression that the classification of results into such two distinct categories would in general be straightforward. On the contrary, however, the *strength* of the recorded incompatibility often *varies* very much in various species, and sometimes even between individual genotypes or/and according to environmental conditions. In addition, different writers have different opinions, when a plant should be called self-incompatible: even a 'clear' (say one-third) diminution in seed set after self-pollination may have been considered sufficient. Therefore, in many species, a simple classification of individual cross results into '+' or '-' ones, would be far from decisive. Meanwhile, in a species like arctic bramble, with its usually very low seed set in selfings, such a "truncated" classification might be applicable without quite frequent errors. Therefore, in this first step of the study, it was postulated that cross results will be *deterministic*, not stochastic. That is, they are postulated to be error-free, giving only right and distinctive '+' or '-' results.

#### 3.2.1. Determining equivalence classes via crosses

Provided no tester clones of known s-genotypes are available, a complete classification of n ramets into equivalence classes could be acquired by a complete diallel system, ie. crossing each of the ramets with each other. To reveal the possible self-fertile clones, also selfings should be included. That is, a complete diallel cross with these n ramets should be exercised, ie.  $n^2$  crosses. Dropping out selfings and reciprocal crosses, ie. making a half-diallel without selfings, would still leave us with  $n(n-1)/2$  crosses to be made, which would not include control crosses and unsuccessful ones. With the number of ramets at present study, this number of crosses would be too large to put into practice. For instance if  $n = 47$ , as in Fig. 13, then altogether 1081 'net' crosses would be needed.

In the present inquiry, 30 shoots were sampled from a population, resulting in a requirement of 435 'net' crosses per population. Allowing for control crosses and un-

successful ones, that figure must be more than doubled — say 1000 crosses per population. Since 29 populations were chosen for study, at least a total number of 29 000 crosses would be needed in subdividing these populations into their equivalence classes. Respectively, using full diallels, 900 'net' crosses per population would be needed, resulting in a total need of at least about 58 000 crosses. According to pilot studies, about 5 crosses per hour could be done by a single person in ideal conditions in June, and much less in July. Hence, much more than 12 000 hours of crossing work — ie. 1500 days, 8 'net' hours of work in each — would have been required. This is not practical. Therefore, other methods have to be sought.

#### 3.2.2. Dynamically optimized crossing

One way of saving in the total amount of crosses needed, would be to divide the total task into a sequence of cycles. Then, the results in each cycle can be recorded and utilized in the planning of the next cycle, thus progressively reducing the part of the population yet to be studied. Thus, the crossing work will be dynamically optimized.

In principle, the greatest saving in the number of crosses to be made, could be achieved, if only one cross (with its control cross) were made at a time. That is, after making a cross, we would wait until its results were recorded and analyzed, before performing the next one. Then, all the previous information would be available in designing each new cross. In practice, however, that will be impossible, bearing in mind that in arctic bramble, a time lag from 10 to 14 days is necessary before the results of a cross can be recorded. Thus, an extreme saving in the numbers of crosses would mean an untenable wasting of time.

Therefore, a compromise between effort and time is necessary. On each day, as *many* crosses as technically possible, are made. Results are recorded and their information analyzed daily. Hence, the guidelines concerning the new crosses to be performed will not be based on the results of all the previous crosses but on all the crosses made up to about 10 days beforehand. This lag will to a some extent diminish the effectiveness for the planning the new crosses. However, the prolific period of arctic bramble flowering in June can now be fully utilized in the crossing work.

##### 3.2.2.1. Gain in efficiency

To obtain a rough picture of the *efficiency* of such a dynamical optimization, we can consider a somewhat simplified version. That is, a guided series of crosses is made in each cycle, their results being recorded and analyzed before planning the next cycle of crosses. In each cycle, one of the ramets is chosen as a universal pollen parent, and crossed with all unclassified ramets still remaining. By recording the results, all ramets giving a negative result with the chosen pollen parental ramet, will be classified into the proper equivalence class and removed from later cycles (Fig. 9). Thus, one equivalence class can be identified per cycle.

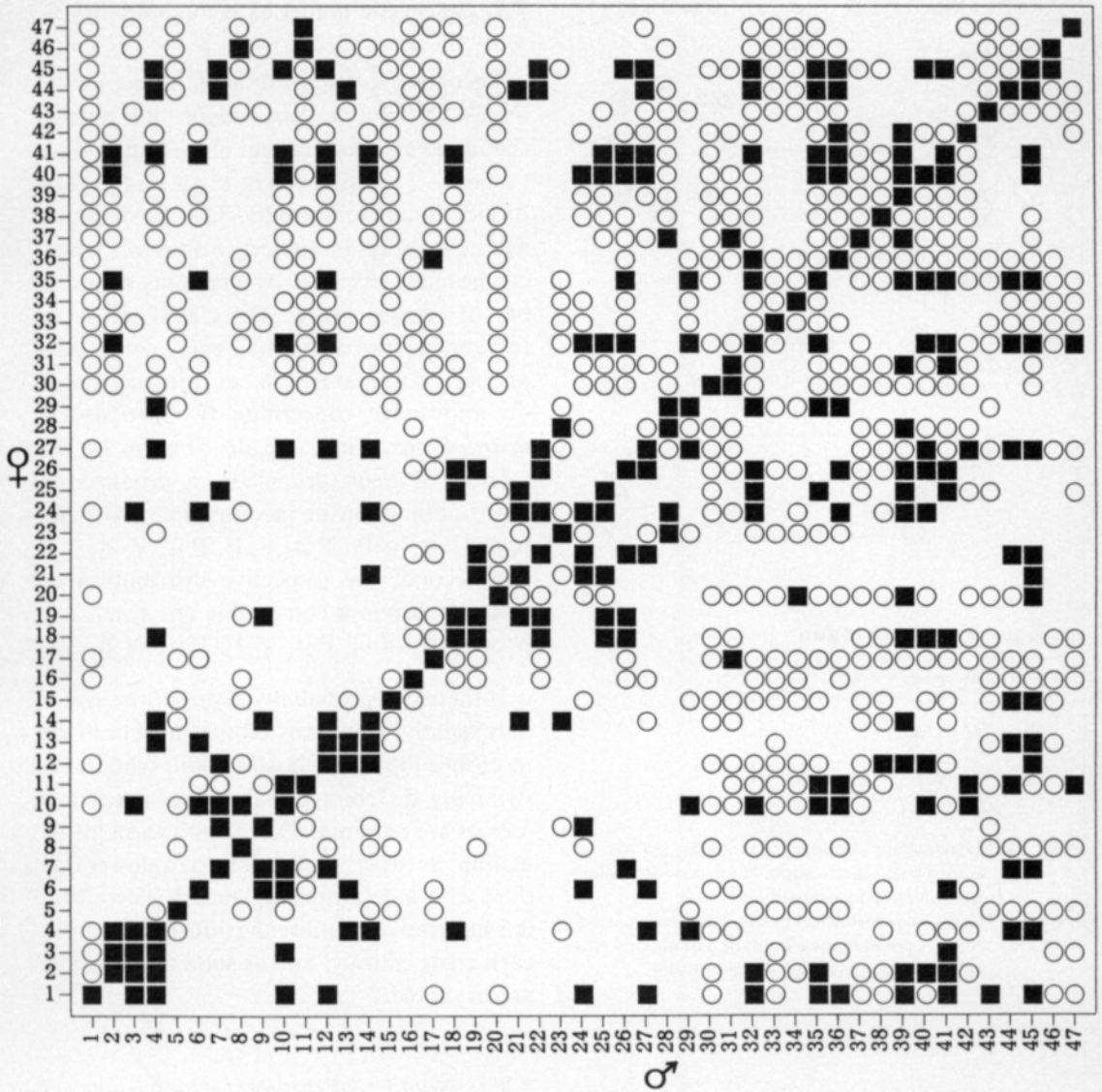


Fig. 13. A (partial) diallel cross between 47 hexaploid *Rorippa sylvestris* (L.) Bess. clones from different locations in N America.

■ = negative result (no seeds were produced),

○ = positive result.

(After MULLIGAN and MUNRO 1984; modified).

The number of remaining ramets diminishes with each cycle, and thereby the crossing algorithm converges quickly. The number of crosses needed, say  $K$ , in a thorough subdivision into equivalence classes, will be

$$(6) \quad K(N, M, \vec{n}; \delta) = -M + \sum_{j=1}^M j \cdot n_{\delta(j)}$$

The notations used in the expression are:  $N$  = number of ramets,  $M$  = number of equivalence classes,  $\vec{n} = (n_1, n_2, \dots, n_M)$  is the vector representing the numbers of

ramets in each equivalence class, listed in descending order, i.e.  $n_i \geq n_{i+1}$  and  $\sum_i n_i = N$ . Furthermore,  $\delta$  denotes a vector of length  $M$ , representing the order in which the equivalence classes are being found. For instance, if the  $i$ th element of the vector  $\delta$  is five, i.e.  $\delta(i) = 5$ , this means that equivalence class number five will be identified on the  $i$ th crossing cycle.

The actual number of crosses needed, is determined by the order, in which the equivalence classes happen to be identified. Hence,  $K$  is a random variable. Expression (6) shows that the number of crosses required would be at

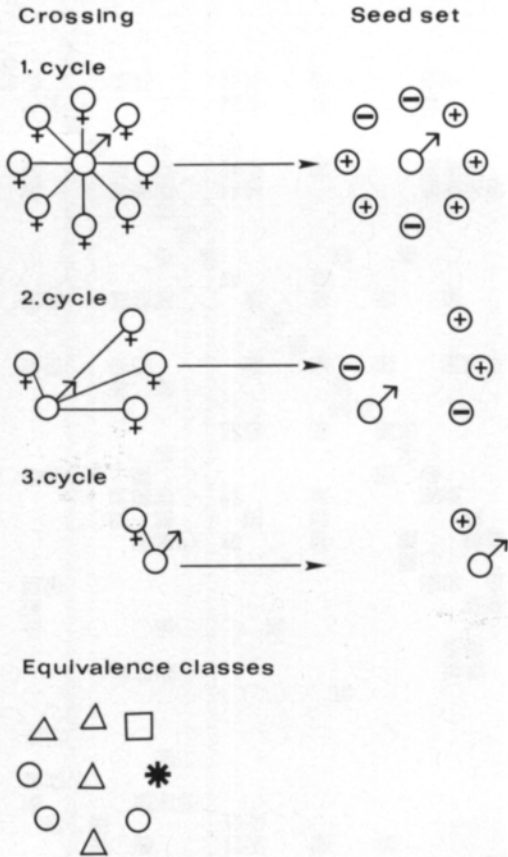


Fig. 9. Dynamically optimized crossing system for equivalence classification, based on a deterministic model of incompatibility.

Crossing is performed in sequential cycles, which renders possible the elimination of already classified ramets from subsequent cycles.

a minimum, if the classes were identified in the order of descending size. And vice versa, unluckily finding the smallest equivalence classes first and the largest ones last, would result in a maximum number of required crosses. However, the order to be realized will be a matter of chance. Fortunately though, the greatest classes (with many ramets) tend to be identified earlier than smaller ones, since one of their ramets is more likely chosen to be the universal pollen parent of the cycle. Hence, in a population with very uneven frequencies of ramets in their equivalence classes, almost a minimum number of crosses will usually suffice (Tab. 4).

When the number of ramets to be studied is, say, at least 30, the benefit given by the optimization will generally be great even in the worst cases, especially if the population contains only few or very uneven equivalence classes. To summarize, by utilizing dynamic optimization, only a few per cent of the crossing work needed in a half-diallel, would be required (Tab. 4).

### 3.3. Stochastic model of incompatibility

Recognizing that the number of seeds set in a cross will be a *random variable*, it is noticed that as a rule, no clear-cut classification into '+' or '-' reaction can be made on the basis of individual cross results. The classification will now always be subject to an *error*, which can be made less likely by increasing the number of crosses made. The classification of two ramets into a common equivalence class or into two separate classes, must be based on knowledge concerning two *probability distributions*. First, the distribution of the number of seeds (drupelets),  $\tau$ , produced in a self-pollination or incompatible cross, ie.  $\bar{p}_s \equiv (P\{\tau_s = 0\}, P\{\tau_s = 1\}, P\{\tau_s = 2\}, \dots)$ , and second, the respective distribution of seed number in a compatible cross, ie.  $\bar{p}_c \equiv (P\{\tau_c = 0\}, P\{\tau_c = 1\}, P\{\tau_c = 2\}, \dots)$ .

If these two probability distributions differ only slightly, very many crosses must be made to ensure the right classification, while with strikingly different distributions, much less crosses are required. The classification information, however, accumulates at a slower rate than with a deterministic model. Therefore, it is imperative to utilize the information from each cross entirely; and as soon as its results are available.

#### 3.3.1. Number of different subdivisions into equivalence classes

With regard to a collection of  $N$  ramets, there will *a priori* be

$$(13) \quad S(N, M) = \frac{1}{M!} \cdot \sum_{i=0}^M (-1)^i \binom{M}{i} (M-i)^N$$

different possibilities of subdividing it into  $M$  equivalence classes (STEINHAUSEN and LANGER 1977). This quantity is called a *Stirling number of the second kind* (cf. 2.2.1.). Summing over possible numbers of classes, ie.  $M = 1, 2, \dots, N$ , will yield the *total number of different subdivisions*, ordinal numbering of classes disregarded,

$$(14) \quad B_N = \sum_{M=1}^N S(N, M).$$



Table 4. Relative amount of crossing labour required, when a dynamically optimized crossing schedule is utilized<sup>5</sup>.

No of equivalence classes (M)	No of ramets to be studied (N)				
	5	10	30	50	100
Minimal required amount <sup>1</sup>					
1	.40	.20	.07	.04	.02
2	.40	.20	.07	.04	.02
3	.50	.22	.07	.04	.02
4	.70	.27	.07	.04	.02
5	1.00	.33	.08	.05	.02
10		1.00	.15	.07	.03
Average amount (extremely uneven classes) <sup>2</sup>					
1	.40	.20	.07	.04	.02
2	.46	.22	.07	.04	.02
3	.60	.26	.07	.04	.02
4	.80	.32	.08	.05	.02
5	1.00	.40	.09	.05	.02
10		1.00	.17	.08	.03
Average amount (extremely even classes) <sup>3</sup>					
1	.40	.20	.07	.04	.02
2		.29	.10	.06	.03
3			.13		
4					.05
5	1.00	.56	.20	.12	.06
10		1.00	.36	.22	.11
Maximal required amount <sup>4</sup>					
1	.40	.20	.07	.04	.02
2	.70	.38	.13	.08	.04
3	.90	.53	.19	.12	.06
4	1.00	.67	.25	.16	.08
5	1.00	.78	.31	.19	.10
10		1.00	.56	.36	.19

<sup>1</sup> See expression (7) in App. H

<sup>2</sup> » » (11) » » »

<sup>3</sup> » » (12) » » »

<sup>4</sup> » » (8) » » »

<sup>5</sup> Expressed in relation to the number of crosses in an ordinary half-diallel without selfings. A deterministic model of incompatibility is postulated.

This quantity is often called a *Bell number*. Its value increases rapidly with N. For example, 10 ramets could a priori be subdivided in  $1.16 \times 10^5$  different ways, 15 ramets in  $1.38 \times 10^9$ , and 30 ramets in  $8.47 \times 10^{23}$  different ways (STEINHAUSEN and LANGER 1977).

Thus, except for collections with only a few ramets, one will have a huge number of different subdivisional hypotheses to deal with. Therefore, among all these hypotheses, it is generally not a simple task to find out, which particular hypothesis is the one that is true.

### 3.3.2. Probability of the actual data on the basis of a hypothesis $H_i$

On the basis of actual cross results, the *right hypothesis* can be selected — or to express it more rigorously, *estimated*. For that purpose, a likelihood connected to each hypothesis is calculated. That is, for each hypothesis  $H_i$  ( $i = 1, 2, \dots, B_N$ ), a probability  $L_i \equiv P_{Hi}$  {actual cross results} is acquired — ie. the probability that (*supposing  $H_i$  holds*) when

performing the series of crosses actually made, cross results identical to the ones actually recorded, would arise. Let  $\tilde{M}$  be the matrix representing the actual cross results recorded. Disregarding the direction in crossing, ie. lumping together reciprocal crosses,

there will be  $\binom{N}{2} = N(N-1)/2$  different pairwise crosses possible among  $N$  ramets. Each different cross is represented by a separate row in  $\tilde{M}$ . A column in the matrix will represent the number of drupelets on a berry, extending from 0 to, say, 70 drupelets. Hence, an element  $m_{k,l}$  in the matrix ( $k = 1, 2, \dots, N(N-1)/2$ ;  $l = 0, 1, \dots, 70$ ), represents the *number of berries* containing exactly  $l$  drupelets, when all crosses between a certain pair of ramets (ie. crosses with an index number  $k$ ) are pooled.

On the basis of a hypothesis  $H_i$ , which describes the equivalence class subdivision of the  $N$  ramets, all the cross results in  $\tilde{M}$  can be further pooled, into no more than two rows. That is, results of all crosses *within* a common equivalence class can be pooled, since they all represent a common probability distribution, ie.  $\bar{p}_s$ . Respectively, all crosses *between* two different equivalence classes, represent  $\bar{p}_c$  and can be pooled into another row. Hence, considering a hypothesis  $H_i$ , the matrix  $\tilde{M}$  of actual results is reduced to *two vectors*, denoted by  $\bar{m}_{is}$  and  $\bar{m}_{ic}$ . Their elements, eg.  $m_{is}(l)$  represent the number of fruits with  $l$  drupelets ( $l = 0, 1, 2, \dots$ ), produced in crosses of the required type — either ‘self or incompatible’ or ‘compatible’. Therefore, *provided  $H_i$  holds*,

$$(15) \quad L_i \equiv P_{H_i} \{\text{results}\} = P_{H_i} \{\bar{m}_{is} \text{ and } \bar{m}_{ic}\}.$$

Due to the independency of crosses, this can be transformed into

$$(16) \quad L_i = P_{H_i} \{\bar{m}_{is}\} \cdot P_{H_i} \{\bar{m}_{ic}\}.$$

It should be noted that each of the probabilities, while representing a repeated (Bernoulli) trial with many (say 71) result categor-

ies, can be acquired from a multinomial distribution. Hence, we finally attain

$$(17) \quad L_i \equiv P_{H_i} \{\text{results}\} \\ = C_i \cdot \prod_l p_s(l)^{m_{is}(l)} \cdot \prod_l p_c(l)^{m_{ic}(l)},$$

$$\text{where } C_i \equiv \frac{m_{is}!}{\prod_l m_{is}(l)!} \cdot \frac{m_{ic}!}{\prod_l m_{ic}(l)!},$$

$$\text{and } m_{is} = \sum_l m_{is}(l)$$

is the total number of (net) ‘self’ type crosses, while  $m_{ic} = \sum_l m_{ic}(l)$  is the total number of ‘compatible’ crosses, postulating  $H_i$  holds. Furthermore, as already defined above,  $p_s(l)$  symbolizes the probability of acquiring exactly  $l$  seeds in a self-pollination or incompatible cross-pollination, and  $p_c(l)$  symbolizes the respective probability in a compatible cross.

In actual fact, likelihood is not defined uniquely, but it can be any function proportional to the probability or probability density in question. Hence, the *constant* coefficient in its defining expression can be chosen arbitrarily. Thus, usually in applications, the multinomial coefficient can be left out from the probability density expression, yielding a simpler and thus more convenient function to be evaluated. In expression (17), however, the coefficient  $C_i$  will *not* be a *constant*, but its value will depend on our parameter under optimization. Namely, in our discrete case, the parameter being considered will be  $i$ , that is the ordinal number of the hypothesis. Therefore, in principle, the term  $C_i$  *cannot* be left out (cf., however, the ‘‘short-hand’’ version, App. I).

### 3.3.3. Estimation of the best hypothesis

Applying the *principle of maximum likelihood* (see eg. LINDGREN 1976, PAHKINEN 1981), the actual cross results are considered to give most support to the hypothesis with the highest L-value. Therefore, using (17), a likelihood value is calculated for each hypothe-

esis  $H_i$  ( $i = 1, 2, \dots, B_N$ ). The hypothesis with the highest L-value can then be selected. This hypothesis is considered to be the best, supposedly true one, as *estimated* on the basis of the (random) outcome of a series of actual crosses.

An actual comparison of the acceptance order of two hypotheses, may be made by considering the *ratio* of their likelihoods or the logarithm of this ratio (called the *support function* eg. by EDWARDS 1972). The hypothesis which has obtained maximum support, is regarded by PAHKINEN (1981) to represent the *least local uncertainty*, in the sense of RENYI (1970). Expressed in more general terms, which can be used in connection with clustering analysis (see KORHONEN 1979), a likelihood value or a ratio of two likelihood values can be employed as a criterion function in clustering. The desired grouping is considered to be the one corresponding to the *extreme value* of the *criterion function*. If likelihood is used as a criterion function, the extreme value to be searched will be a maximum, while if a likelihood ratio is in use, then a minimum value may be required instead (see 3.3.4.).

### 3.3.3.1. Likelihood ratio of two arbitrary subdivisions, $H_i$ and $H_j$

Applying expression (17), we can derive the ratio of the likelihoods of two different hypotheses,  $H_i$  and  $H_j$ :

$$(18) \quad L_i/L_j \equiv P_{H_i} \{\text{results}\} / P_{H_j} \{\text{results}\} \\ = \frac{C_i \cdot \prod_l p_s(l)^{m_{is}(l)} \cdot \prod_l p_c(l)^{m_{ic}(l)}}{C_j \cdot \prod_l p_s(l)^{m_{js}(l)} \cdot \prod_l p_c(l)^{m_{jc}(l)}} \\ = \frac{C_i}{C_j} \cdot \prod_l \{p_s(l)^{[m_{is}(l) - m_{js}(l)]} \cdot p_c(l)^{[m_{ic}(l) - m_{jc}(l)]}\}.$$

When changing from the hypothesis  $H_i$  into  $H_j$ , certain of the crosses made will change in their interpretation from a 'self' to a 'compatible' category. That is, they will *not* be counted into  $m_{js}(l)$ , but into  $m_{jc}(l)$ . On the other hand, also some of the cross results may change their type from a 'compatible' to a 'self' category. Such

cross results, which were before counted into  $m_{ic}(l)$ , will now, under  $H_j$ , be counted into  $m_{js}(l)$ . The difference  $m_{is}(l) - m_{js}(l)$ , will contain only the  $l$ -drupelet results with such a change in category, ie.  $m_{is}(l) - m_{js}(l) = \#$  [crosses changing type from 'self' in  $H_i$  to 'compatible' in  $H_j$ ] with a result of  $l$  drupelets]  $-$   $\#$  [crosses changing type from 'compatible' in  $H_i$  to 'self' in  $H_j$ ] with a result of  $l$  drupelets], where  $\# \{ \}$  denotes the *number of elements* in a set. This expression can be notated in a still more compact manner as follows  $m_{is}(l) - m_{js}(l) = \#$  [is-jc |  $l$  drupl.]  $-$   $\#$  [ic-js |  $l$  drupl.]. Similarly,  $m_{ic}(l) - m_{jc}(l) = \#$  [ic-js |  $l$ ]  $-$   $\#$  [is-jc |  $l$ ]  $= -$ [ $m_{is}(l) - m_{js}(l)$ ].

Hence, expression (18) can be developed into

$$(19) \quad \frac{L_i}{L_j} \equiv \frac{P_{H_i} \{\text{results}\}}{P_{H_j} \{\text{results}\}} = \frac{C_i}{C_j} \cdot \prod_l \left\{ \frac{p_s(l)}{p_c(l)} \right\}^{(\# \text{ [is-jc | } l \text{]} - \# \text{ [ic-js | } l \text{]})} \\ \equiv C_{ij} \cdot T_{ij},$$

where  $C_{ij}$  denotes the 'coefficient part' and  $T_{ij}$  the 'probability part'.

Next, we will consider the coefficient part. Almost directly from the definition in (17), after only a slight re-ordering, we will obtain,

$$(20) \quad C_{ij} \equiv \frac{C_i}{C_j} = \frac{m_{is}!}{m_{js}!} \cdot \frac{m_{ic}!}{m_{jc}!} \cdot \prod_l \left\{ \frac{m_{js}(l)!}{m_{is}(l)!} \cdot \frac{m_{jc}(l)!}{m_{ic}(l)!} \right\}.$$

### 3.3.4. An extended single move method for finding a locally optimal hypothesis

As was shown above (3.3.1.), except for quite a few ramets, the number of different classifications will be far too large to be inspected enumeratively, ie. calculating a criterion function value (17) for each classification consecutively. Furthermore, our problem deals not with analytic, ie. continuous and differentiable functions, but with a discrete set of values. Accordingly, no general optimization theory for finding an *absolute* maximum point was available.

Thus, an *algorithm* was formulated, which — when started from any initial, proposed

subdivision — should finally yield a locally optimal subdivision into equivalence classes, corresponding to a *local* maximum value of the criterion function. Following the nomenclature by FRIEDMAN and RUBIN (1967), such a point (or subdivision) could be called a *single move local maximum point* for the criterion.

In short, such a single-move algorithm consists of following steps. Starting from a hypothesis  $H_i$ , we take into consideration all hypotheses “surrounding it at a distance of one move”, i.e. every hypothesis  $H_j$  such that it can be attained via *changing the classification* of no more than a *single ramet* in  $H_i$ . The change will be either a) removing one ramet from an equivalence class, without leaving the class empty, and transferring the ramet into another, already existing class, or b) amalgamating a one-ramet class into another class, or c) removing a ramet from an equivalence class and letting the ramet in question constitute a new equivalence class. The likelihood of each possible “surrounding” hypothesis  $H_j$ , is compared separately with the likelihood of  $H_i$ , by forming their likelihood ratio (18 or 19, 20). If the cross result data gives more support to any one  $H_j$  than to  $H_i$  (i.e.  $L_i/L_j < 1$ ), then the hypothesis  $H_i$  is rejected and *replaced* by one of the surrounding hypotheses  $H_j$  — i.e. by the one with the greatest support against  $H_i$  (that is, with the lowest  $L_i/L_j$  value), say  $H_{j_0}$ .

Then a new cycle of the algorithm is begun, taking  $H_{j_0}$  as the reference basis, i.e. as a new starting hypothesis. The algorithm stops, when no hypothesis  $H_m$  — at the prescribed one-move distance — can replace the latest ‘starting’ hypothesis, say  $H_{j_k}$ , i.e. when  $L_{j_k}/L_m \geq 1$  for every possible  $m$  value. The resulting hypothesis,  $H_{j_k}$ , represents the *single move locally optimal subdivision* of the ramets into equivalence classes.

This algorithm is, however, an ‘extended’ one (see later), in the sense that it occasionally contains also moves of more than one ramet at a time — therefore it is able to inspect “more far” from the initial point than a purely single move method would do, in searching for a higher valued point.

For a detailed description of the single move method, see App. I.

### 3.3.4.1. Basic formulas in the algorithm

All calculations in the present study have been performed *applying*  $T_{ij}$  (the *probability part* in expression 19) as the *criterion* function. That is, in choosing between two hypotheses, the ‘short-hand’ version (see App. I, chapter 4) is adopted. Reasons for this choice are considered in App. I. First, in  $T_{ij}$  the probability of an *actual result* seems to be represented in a more *adequate* way than in the ‘general’ criterion  $C_{ij} \cdot T_{ij}$ . In the latter function, the coefficient part emerges, since non-occurred cases have also been (in an improper way) included in the ‘actual result’. Secondly, the *classificatory power* of  $T_{ij}$  seems consistently to be quite high, while a clear evidence for this could not be given in regard with the ‘general’ criterion function. Furthermore, the *calculation effort* is much smaller, when only  $T_{ij}$  is used.

Denoting a cross  $r \times r_0$  as well as its reciprocal  $r_0 \times r$  by an index number  $k_r$ , the number # {is → jc | l} can be expressed as

$$\# \{is \rightarrow jc | l\} = \sum_{\substack{r \in e_i \\ r \neq r_0}} m_{k_{r,l}} \equiv \sum_{\substack{r \in e_i \\ r \neq r_0}} \# \{r \times r_0 \text{ or } r_0 \times r | l\}.$$

This is shown in App. I, chapter 1. Hence, the criterion function  $T_{ij}$  (see expression 19), respecting a move of a ramet  $r_0$  from a ‘previous’ equivalence class  $e_i$  (in hypothesis  $H_i$ ) into a ‘forthcoming’ equivalence class  $e_j$  (in hypothesis  $H_j$ ), is presented as

$$(41) \quad T_{ij} = \prod_l \left\{ \frac{p_s(l)}{p_c(l)} \right\}^{\left[ \sum_{\substack{r \in e_i \\ r \neq r_0}} m_{k_{r,l}} - \sum_{\substack{r \in e_j \\ r \neq r_0}} m_{k_{r,l}} \right]},$$

or its two special cases as

$$(42) \quad T_{ij} = \prod_l \left\{ \frac{p_s(l)}{p_c(l)} \right\}^{\left[ - \sum_{\substack{r \in e_j \\ r \neq r_0}} m_{k_{r,l}} \right]},$$

when  $e_i = \{r_o\}$ , and

$$(43) \quad T_{ij} = \prod_l \left( \frac{p_s(l)}{p_c(l)} \right)^{\left[ \sum_{\substack{r \in e_i \\ r \neq r_o}} m_{kr,l} \right]}$$

when  $e_j = \{r_o\}$ .

If  $T_{ij} < 1$ , then hypothesis  $H_j$  is supported against hypothesis  $H_i$ .

### 3.3.4.2. Extension: amalgamation of two equivalence classes

The single move method is extended by including occasional steps of merging two equivalence classes, each of which contain more than one ramet. This turned out to be manageable in practice, since the fall-out of the coefficient part will greatly reduce the calculation effort required. Hence, whenever the single-move algorithm would stop, the search for a still more supported hypothesis is continued, by inspecting hypotheses further than a distance of one ramet move from the starting hypothesis  $H_i$  of the cycle. This additional search is, however, confined only to such hypotheses  $H_j$ , which can be acquired via merging a single pair of equivalence classes in  $H_i$ .

If any of these 'fusion' hypotheses  $H_j$  is supported against  $H_i$ , ie. if  $T_{ij} < 1$  in the expression (44) (see below) for at least one  $H_j$ , then the most supported one of these is chosen as a new starting hypothesis, and a new cycle of the single-move algorithm is started.

Directly from expression (19), we obtain

$$(44) \quad T_{ij} \equiv \prod_l \left( \frac{p_s(l)}{p_c(l)} \right)^{\left( \# [is-jc | l] - \# [ic-js | l] \right)}$$

Let two equivalence classes, say  $e_1$  and  $e_2$ , in  $H_i$  be fused together to constitute an equivalence class  $e_{12}$  in  $H_j$ . Then only crosses changing in type with a change from  $H_i$  to  $H_j$ , are those made between the classes  $e_1$  and  $e_2$  — they change from 'compatible' to 'self' type — while all crosses within  $e_1$  and  $e_2$  respectively, remain 'self' in type. Hence,  $\# [is-jc | l] = 0$  and

$$\# [ic-js | l] = \sum_{\substack{r_1 \in e_1 \\ r_u \in e_2}} \# \{r_1 \times r_u \text{ or } r_u \times r_1 | l\},$$

which is denoted by  $\sum_{\substack{r_1 \in e_1 \\ r_u \in e_2}} m_{k_{tu},l}$ ,

where the index  $k_{tu}$  points to a cross  $r_1 \times r_u$  or its reciprocal  $r_u \times r_1$ .

Hence, the expression (44) can be modified into

$$(45) \quad T_{ij} = \prod_l \left( \frac{p_s(l)}{p_c(l)} \right)^{\left[ - \sum_{\substack{r_1 \in e_1 \\ r_u \in e_2}} m_{k_{tu},l} \right]}$$

which can directly be used in practical calculations, for an 'amalgamation step' between sequences of one ramet move cycles, in the extended algorithm.

### 3.3.4.3. Comfortable measures

For computational and illustrative purposes, as well as aiming at an economy in notations, it is practical to reorganize the treatment presented above. Hence, all seed numbers data concerning a certain cross, say  $r_1 \times r_2$  (including its reciprocal form), is concentrated into a single one fundamental item, to be called 'similarity' of ramets  $r_1$  and  $r_2$ , and denoted by  $S(r_1, r_2)$ . Namely,

$$(46) \quad S(r_1, r_2) \equiv \log \prod_{r_1 \times r_2} \left( \frac{p_s(l_{r_1 \times r_2})}{p_c(l_{r_1 \times r_2})} \right) \\ = \sum_{r_1 \times r_2} \log \left( \frac{p_s(l_{r_1 \times r_2})}{p_c(l_{r_1 \times r_2})} \right) \\ = \sum_{r_1 \times r_2} [\log p_s(l_{r_1 \times r_2}) - \log p_c(l_{r_1 \times r_2})].$$

Utilizing this term, the basic formula (41) of one ramet move algorithm, can be reformulated into

$$(41') \quad \log T_{ij} = \sum_{\substack{r \in e_i \\ r \neq r_o}} S(r_o, r) - \sum_{\substack{r \in e_j \\ r \neq r_o}} S(r_o, r),$$

with the two special cases

$$(42') \quad \log T_{ij} = - \sum_{\substack{r \in e_j \\ r \neq r_o}} S(r_o, r),$$

when  $e_i = \{r_o\}$ , and

$$(43') \log T_{ij} = \sum_{\substack{r \in e_i \\ r \neq r_0}} S(r_0, r),$$

when  $e_j = \{r_0\}$ .

Similarly for expression (45), which regards the amalgamation of equivalence classes  $e_1$  and  $e_2$ , a more convenient form is obtained, using 'similarity':

$$(45') \log T_{ij} = - \sum_{\substack{r_i \in e_1 \\ r_u \in e_2}} S(r_i, r_u),$$

where  $H_j$  is the same as  $H_i$  but with equivalence classes  $e_1$  and  $e_2$  merged.

It is straightforward to introduce respective higher level terms, ie. 'similarity' of a ramet  $r_0$  to an equivalence class, say  $e_i$ , defined as

$$(47) S(r_0, e_i) \equiv \sum_{\substack{r \in e_i \\ r \neq r_0}} S(r_0, r),$$

and 'similarity' of two equivalence classes, say  $e_i$  and  $e_j$ , defined as

$$(48) S(e_i, e_j) \equiv \sum_{r \in e_i} S(r, e_j).$$

Applying these definitions, the formulas above are simplified further, and we obtain for a *one ramet move* (considering  $H_i$ :  $r_0 \in e_i$  versus  $H_j$ :  $r_0 \in e_j$ ),

$$(41'') \log T_{ij} = S(r_0, e_i) - S(r_0, e_j), \text{ or}$$

$$(42'') \log T_{ij} = -S(r_0, e_j), \text{ when } e_i = \{r_0\}, \text{ or}$$

$$(43'') \log T_{ij} = S(r_0, e_i), \text{ when } e_j = \{r_0\},$$

and for *amalgamation of classes  $e_1$  and  $e_2$*  (considering  $H_i$ :  $e_1$  and  $e_2$  are separate equivalence classes versus  $H_j$ :  $e_1$  and  $e_2$  merged constitutes an equivalence class  $e_{12} = e_1 \cup e_2$ ),

$$(45'') \log T_{ij} = -S(e_1, e_2) = - \sum_{r \in e_1} S(r, e_2).$$

Regarding the 'similarity' definitions, it should be noted that they are here made on the *logarithmic* basis.

Hence,  $H_j$  will be supported against  $H_i$ , if and only if  $\log T_{ij} < 0$ . However, the 'similarities' could as well have been defined without the use of logarithms — the sums and differences would then have been products and quotients, respectively.

On the basis of the 'similarity' concept, additional illustrative terms are introduced later (see 3.3.5.5 and 3.3.5.7).

### 3.3.5. Choice of the most informative forthcoming crosses

In a population with  $n$  ramets, there will be  $n(n-1)/2$  different crosses possible. At least within the deterministic model (see Chapter 3.2), a *random choice* from all possible crosses would be *highly inefficient*. Therefore, it would be better to utilize the *information* that accumulates together with the cross results that are recorded, for an optimal design of new crosses.

#### 3.3.5.1. The effect of a single extra cross on likelihood ratio

Assume that, after completing the previous ( $n^{\text{th}}$ ) cycle of crosses, we have made a *single extra cross*, say  $r_1 \times r_2$ . Suppose that this extra cross has yielded a result of  $l_{r_1 \times r_2}$  seeds.

From the 'general' likelihood ratio expression (19), we obtain the logarithmic version of the criterion function,

$$(49) \log T_{ij} = \sum_j [ \# \{is \rightarrow jc \mid l\} - \# \{ic \rightarrow js \mid l\} ] \cdot \log \left[ \frac{P_s(l)}{P_c(l)} \right].$$

Applying this formula, the *change* in criterion function value, caused by the crosses made in the  $(n+1)^{\text{th}}$  cycle, are expressed as

$$(50) \delta^{(n)} \log T_{ij} \equiv \log T_{ij}^{(n+1)} - \log T_{ij}^{(n)} \\ = \sum_j [ \#^{(n+1)} \{is \rightarrow jc \mid l\} - \#^{(n)} \{is \rightarrow jc \mid l\} \\ - \#^{(n+1)} \{ic \rightarrow js \mid l\} + \#^{(n)} \{ic \rightarrow js \mid l\} ] \cdot \log \left[ \frac{P_s(l)}{P_c(l)} \right].$$

Since the  $(n+1)^{\text{th}}$  crossing cycle consists of only a single cross,  $r_1 \times r_2$ , we obtain

$$\delta^{(n)} \log T_{ij} = [ \#^{(n+1)} [is \rightarrow jc \mid l_{r_1 \times r_2}] - \#^{(n)} [is \rightarrow jc \mid l_{r_1 \times r_2}] \\ - \#^{(n+1)} [ic \rightarrow js \mid l_{r_1 \times r_2}] + \#^{(n)} [ic \rightarrow js \mid l_{r_1 \times r_2}] ] \cdot \log \frac{p_s(l_{r_1 \times r_2})}{p_c(l_{r_1 \times r_2})}$$

Hence, the effect of a single extra cross  $r_1 \times r_2$  is

$$(51) \quad \delta^{(n)} \log T_{ij} = \begin{cases} \log \{p_s(l_{r_1 \times r_2})/p_c(l_{r_1 \times r_2})\} \\ \text{if } r_1 \times r_2 \in \{is \rightarrow jc\}, \text{ or} \\ -\log \{p_s(l_{r_1 \times r_2})/p_c(l_{r_1 \times r_2})\} \\ \text{if } r_1 \times r_2 \in \{ic \rightarrow js\}, \text{ or} \\ 0 \text{ otherwise (ie. if the} \\ \text{relation between } r_1 \text{ and} \\ r_2 \text{ is the same} \\ \text{in } H_j \text{ as in } H_i). \end{cases}$$

The contents of this result may become clearer, if presented in a slightly different manner:

#### Theorem I.

A new cross  $r_1 \times r_2$  to be made, will a) not change the support of any hypothesis against any other hypothesis, in which the relation between  $r_1$  and  $r_2$  (ie. them being equivalent or not) is stated in an identical way — totally disregarding, whether both hypotheses are true or both are untrue in this respect; and b) increase the support of any hypothesis, which states the relation between  $r_1$  and  $r_2$  in a true way, against each hypothesis, which states this relation in a false way.

This increase will be by an amount

$$\log \{p_s(l_s)/p_c(l_s)\},$$

if (in real world)  $r_1$  is equivalent to  $r_2$ , while it will be by an amount

$$-\log \{p_s(l_c)/p_c(l_c)\},$$

if  $r_1$  and  $r_2$  are not equivalent. This latter amount of increase will on average be greater than the former one.

Here  $l_s$  denotes a random seed set result in a 'self' type (incompatible) cross, ie. it will be taken from the distribution  $p_s(l)$ , and  $l_c$  denotes a random seed set result in a 'compatible' cross, ie. it will be taken from the distribution  $p_c(l)$ .

— As random variables, each amount of in-

crease described in b), will occasionally attain negative values, which then constitutes a decrease instead. In the great majority of cases, however, they will definitely be positive. Furthermore, it can be proved that on average, the greatest increase in the support of a " $r_1, r_2$ -true" hypothesis against a " $r_1, r_2$ -false" hypothesis, will be achieved in the latter instance, ie. if  $r_1$  and  $r_2$  are non-equivalent.

Namely, regarding a strictly self-incompatible species, the discriminatory power of the quotient function  $p_s(l)/p_c(l)$  will be at its greatest at the modal point (most commonly occurring value) of the 'compatible' distribution  $p_c(l)$ , respecting a minimum point of the quotient function (see Figs 10 and 11). Since, the two distributions,  $p_s(l)$  and  $p_c(l)$ , will have still more dissimilar values there than at the maximum point of the quotient function (near the origin), which represents the modal point of the  $p_s(l)$  distribution. Stated more precisely,

$$1/\min_l \frac{p_s(l)}{p_c(l)} > \max_l \frac{p_s(l)}{p_c(l)}.$$

Therefore, in most instances,

$$1/\frac{p_s(l_c)}{p_c(l_c)} > \frac{p_s(l_s)}{p_c(l_s)}.$$

Hence, a high seed number will indicate a 'compatible' cross far more decisively than a low seed number will indicate a 'self' type cross.

The reason for this is that we seldomly obtain high numbers of seeds in an actual incompatible cross. The few cases which possibly occur, are man made errors (technical mistakes, eg. wrong source of pollen, disordering of labels, error in writing crossing notes, careless isolation of flowers, etc.), or errors by nature (a constantly or occasionally self-fertile clone). Organizing crossing work carefully, and including controls in each step, the frequency of such technical errors can be kept low. On the contrary, in actually compatible crosses, scattered cases with low seed numbers will be recorded moderately often, despite careful control over the effectivity of pollen etc. Other factors which might influence low seed numbers are unnoticed poor quality of the flower of the seed parent — eg. due to a powdery mildew, thrips, damage to the pistils in emasculation, too young (not yet receptive) buds used in a cross — as well as inconsistent effects of a moderate rain, trampling etc.

#### 3.3.5.2. Preferable crosses ...

If the real situation concerning the relation between any two ramets is known, then Theorem I could be used to choose suitable crosses between non-equivalent ramets, rendering the collection of evidence for the (al-

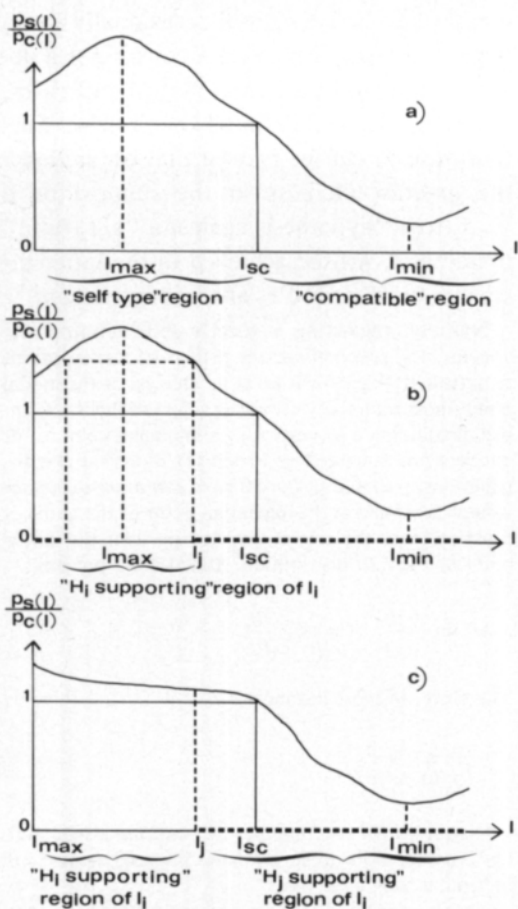


Fig. 10. Quotient function of the two probability distributions,  $p_s(l)$  and  $p_c(l)$ .

a) Subdivision of the seed number axis ( $l$ ) into two regions:  $l < l_{sc}$ , where  $p_s(l) > p_c(l)$ , and  $l > l_{sc}$ , where  $p_s(l) < p_c(l)$ . Seed numbers respecting the maximum ( $l_{max}$ ) and minimum ( $l_{min}$ ) values of the quotient function, have been pointed out.

b) Subdivision of the seed number axis into two regions, "H<sub>i</sub> supporting region" and its complement, "H<sub>j</sub> supporting region". The latter is denoted by a dashed line.

c) Respective subdivision as in b) but in a strictly self-incompatible plant species.

ready known) true hypothesis H<sub>i</sub> maximally efficient. Generally, however, such a *prior* knowledge is *lacking*. All the information concerning the partition of the population into equivalence classes, remains to be acquired from the crosses.

In the primary, unanalyzed form, this information is available in a matrix of pairwise cross results. The first

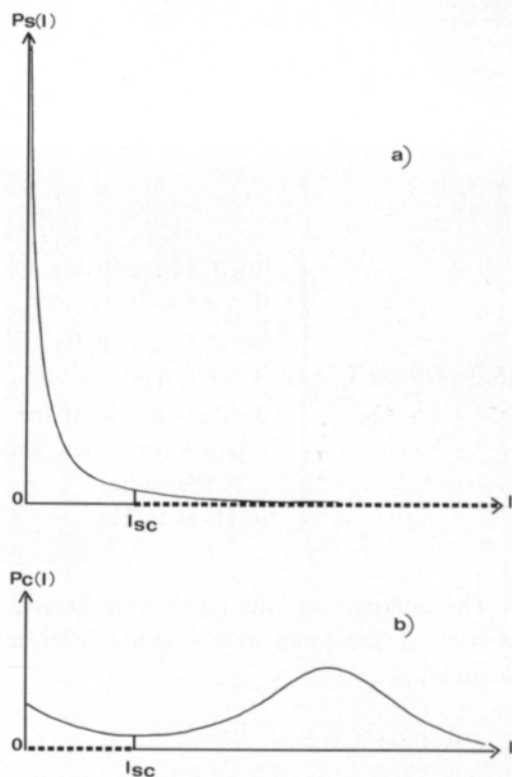


Fig. 11. Plausible types of seed probability distributions in a strictly self-incompatible plant species. a) Incompatible pollinations, b) compatible pollinations.  $l$  = seed number. Critical values are denoted by a dashed line, i.e. "compatible" region in an incompatible pollination, or "self type" region in a compatible pollination (cf. Fig. 10a)

analyzed form of the data is a matrix of pairwise 'similarities' between ramets; in its production, knowledge of the quotient function  $p_s(l)/p_c(l)$  is required. In its highest analyzed form, this accumulated information is transformed into a single move (locally optimal) estimate  $H_0^{(n)}$  of the best hypothesis, accompanied by an ordered sequence of less and less likely hypotheses.

The locally optimal hypothesis acquired after  $n$  crossing cycles,  $H_0^{(n)}$ , will often be *false*, at least during the first cycles of the crossing study. That is, it will not state the same partition into equivalence classes as the true hypothesis H<sub>i</sub> does. If our recursive system of guided experimentation and analysis is to function at all correctly, however, the hypothesis  $H_0^{(n)}$  should become on average *less and less false* during the crossing study. This



means that more and more information will accumulate into  $H_0^{(n)}$ , and of the  $N(N-1)/2$  pairwise relations existing between  $N$  ramets, it will likely state ever *greater and greater proportion* in a *true* way. Hence, if we aim at utilizing the information of previous crosses in an optimal design of new crosses, we *must* — in one way or another — take into consideration  $H_0^{(n)}$ , our present (“ $n^{\text{th}}$  crossing cycle, locally optimal one ramet move”) *estimate* of the best hypothesis.

Therefore, from this point onwards, we will as a rule compare, not any two subdivisional hypotheses, but  $H_0^{(n)}$  to certain other hypotheses, usually to the hypotheses at a *single ramet move distance* from it.

Considering Theorem I from a slightly different angle, and from the point of view of  $H_0^{(n)}$ , we obtain

#### Theorem II.

If we choose the two partners of the extra cross  $r_1 \times r_2$  from

- I. a common equivalence class (according to  $H_0^{(n)}$ ), and if
  - A.  $H_0^{(n)}$  is “ $r_1, r_2$ -true”, then the extra cross will *increase* the support of  $H_0^{(n)}$ , but not maximally;
  - if, on the contrary,
  - B.  $H_0^{(n)}$  is “ $r_1, r_2$ -false”, then the extra cross will *decrease* the support of  $H_0^{(n)}$  *maximally*.

While, if we choose the two partners of the extra cross  $r_1 \times r_2$  from

- II. two *different* equivalence classes (according to  $H_0^{(n)}$ ), and if
  - A.  $H_0^{(n)}$  is “ $r_1, r_2$ -true”, then the extra cross will *increase* the support of  $H_0^{(n)}$  *maximally*;
  - if, on the contrary,
  - B.  $H_0^{(n)}$  is “ $r_1, r_2$ -false”, then the extra cross will *decrease* the support of  $H_0^{(n)}$ , but not maximally.

The increase or decrease described occurs with respect to any hypothesis *opposite* to  $H_0^{(n)}$  concerning the relation between  $r_1$  and  $r_2$ . At the prescribed *single ramet move distance* to  $H_0^{(n)}$ , there is, in case I, as many as  $2 \cdot m_0$  such hypotheses (where the number of equivalence classes in  $H_0^{(n)}$  is denoted by  $m_0$ ), while in case II, only two such hypotheses are found.

Hence when considering  $H_0^{(n)}$  to be not yet totally true, as a rule, the *general strategy* would be to *attack* at its *false points*, in order to abandon it and then *replace* it by a new, *less false* hypothesis. Assuming that this attack is successful, Theorem II tells that by hitting a cross at a false relation *within* an equivalence class (ie. a pair of ramets falsely classified into a common equivalence class), would result in a *maximal* decrease in the support of  $H_0^{(n)}$ , which is the desired result. Therefore, if the false relations could always be pinpointed, then utilizing *procedure I.B.* would be optimal. Since, however, the true situation is not known, some wrong guesses may be made. As a consequence, we will sometimes fall into I.A., when aiming at I.B., thence introducing an increase instead. This increase, however, is not fatal, since it is an increase in support of  $H_0^{(n)}$  and other “ $r_1, r_2$ -true” hypotheses against all “ $r_1, r_2$ -false” ones, while it will not change at all the support of  $H_0^{(n)}$  against these other “ $r_1, r_2$ -true” hypotheses. In the short term, at least, it will however *slow down* the replacement of  $H_0^{(n)}$  by a less false hypothesis. Therefore, the actual efficiency of each of the procedures, I and II, is decisively dependent upon, how successfully under each procedure can right guesses concerning false relations in  $H_0^{(n)}$  be made.

The average efficiency of procedures I and II in different kinds of populations, is studied in App. J. In most cases, procedure I would usually be more profitable, except in uniclassic populations, where procedure II would be more efficient.

#### 3.3.5.3. ... and how to find them

Disregarding whether the crossing partners are taken within a common class or from two different classes, a *random sampling* may prove to be quite *inefficient*. Especially towards the end of the study, it may be difficult to hit at a false relation at all (App. J). Hence, at least when the crossing study proceeds, yielding at *less and less false* locally optimal hypotheses  $H_0^{(n)}$ , a more efficient way of at-

tacking the false relations in  $H_0^{(n)}$  must be found.

Until now, we have utilized only that part of the cross results information, which has accumulated into the *acceptance order* of the partitioned hypotheses concerned. Now it is time to take into consideration that the individual 'similarity' values actually contain detailed information at the level of ramet pairs, and also information of individual ramets with regard to their rivalling equivalence classes. This information can be used in a search for false relations in  $H_0^{(n)}$ .

A 'similarity' of two ramets (see 3.3.4.3, expression 46) actually contains the *likelihood ratio* of their pairwise cross results, taken from a 'self' type versus a 'compatible' distribution. Hence, a positive value would indicate that — judging solely on their pairwise cross results — the particular two ramets will more plausibly be interpreted as equivalent than as non-equivalent ones.

It can also be shown that the 'similarity' of a ramet, say  $r_1$ , to any proposed equivalence class, say  $e_1$ , is *sensitive* to changes in the *constitution* of  $e_1$ . Suppose that the equivalence classes in  $H_0^{(n)}$  are not yet true equivalence classes but *mixtures of ramets* from different true equivalence classes.

Let us denote the true classes by  $et_i$ ,  $i = 1, \dots, m_t$ , and the classes in  $H_0^{(n)}$  by  $e_u$ ,  $u = 1, \dots, m_o$ . In the class  $e_u$ , we will denote by  $e_{ui}$  the particular subclass, which contains ramets from  $et_i$ ; that is,  $e_{ui}$  is the intersection  $e_u \cap et_i$ . Furthermore, let  
 $n_u$  = number of ramets in the class  $e_u$ ,  
 $n_{ui}$  = » » » » » subclass  $e_{ui}$ ,  
 $nt_i$  = » » » » » true class  $et_i$ , where  
 $i = 1, \dots, m_t$  and  $u = 1, \dots, m_o$ .

Thence, any class  $e_u$  in  $H_0^{(n)}$  can be represented as a union of such subclasses,

$$(58) \quad e_u = \bigcup_{i=1}^{m_t} e_{ui} = \bigcup_{i=1}^{m_t} (e_u \cap et_i),$$

where any but one of the intersections may be empty. Then, supposing that  $r_1 \in et_1$ , the following equation is obtained

$$(59) \quad S(r_1, e_u) = \sum_{\substack{r \in e_{u1} \\ r \neq r_1}} S(r_1, r) + \sum_{\substack{r \in e_u \\ r \notin e_{u1}}} S(r_1, r) \\ = \sum_{\substack{r_1 \times r \\ r \in e_{u1} \\ r \neq r_1}} \log \frac{p_s(l_s)}{p_c(l_s)} + \sum_{\substack{r_1 \times r \\ r \in e_u \\ r \notin e_{u1}}} \log \frac{p_s(l_c)}{p_c(l_c)}.$$

Since the quotient function  $p_s(l)/p_c(l)$  is, as a rule, much lower (see Theorem I, Figs 10 and 11), if the number of seeds ( $l$ ) is taken from a 'compatible' than from a 'self' type distribution, this 'similarity' value will in most cases be lower than if all ramets in  $e_u$  were truly equivalent to  $r_1$ . And more generally:

*Lemma I.*

Replacing a  $r_1$ -equivalent ramet in  $e_u$  with a  $r_1$ -non-equivalent one, will on average result in a *decrease* in  $S(r_1, e_u)$ .

From the same formula we can deduce also the following, somewhat vague conclusion.

*Corollary I.*

$S(r_1, e_u) > 0$  as a rule only if a definite majority of the crosses of  $r_1$  within  $e_u$  are with  $r_1$ -equivalent ramets.

Considering procedure I, where both crossing partners are chosen from a common equivalence class (according to  $H_0^{(n)}$ ), we can use the Total Probability Theorem to attain at

$$(60) \quad P_1 \{\text{hit at a false relation}\} \\ = \sum_{\substack{u=1, \dots, m_o \\ i=1, \dots, m_t \\ n_{ui} \geq 2}} P_1 \{\text{hit at a f. rel.} \mid 1^{\text{st}} \text{ partner} \in e_{ui}\} \cdot \\ \cdot P_1 \{1^{\text{st}} \text{ partner} \in e_{ui}\},$$

where  $\sum_{\substack{u=1, \dots, m_o \\ i=1, \dots, m_t \\ n_{ui} \geq 2}} P_1 \{1^{\text{st}} \text{ partner} \in e_{ui}\} = 1$ .

If the *second partner* is chosen *randomly* in the same class (according to  $H_0^{(n)}$ ) from which the *first partner* has been chosen, then

$$(61) P_1 \{ \text{hit at a false rel.} \mid 1^{\text{st}} \text{ partner} \in e_{ui} \} \\ = \frac{(n_u - n_{ui})}{(n_u - 1)},$$

where  $n_u \geq 2$ . Namely, in all cases in which the second partner belongs to another subclass  $e_{uj}$  ( $j \neq i$ ) of  $e_u$ , the first partner is *not* equivalent to the second one, in contrary to what  $H_0^{(n)}$  states. Hence, provided the second partner is chosen randomly in the same class, expression (60) will yield

$$(62) P_1 \{ \text{hit at a false relation} \} \\ = \sum_{\substack{u=1, \dots, m_o \\ i=1, \dots, m_t \\ n_{ui} \geq 2}} \frac{(n_u - n_{ui})}{(n_u - 1)} \\ \cdot P_1 \{ 1^{\text{st}} \text{ partner} \in e_{ui} \}.$$

It should be noted that the coefficient part will be the *greater*, the smaller is the class and the smaller proportion of it is constituted by the subclass. Let us denote by  $I_r$  the version of procedure I, in which the first partner is chosen purely randomly. Under  $I_r$ , (on the contrary to the coefficient part) the probability part in expression (62) will be the *smaller*, when the subclass is smaller. Therefore, regarding the behaviour of the probability part, any change in the direction of the coefficient part — making the probability part great or small *simultaneously* with the coefficient part — will increase the total probability.

Hence, we are searching for a *strategy of choosing the first partner* in such a way that relatively *small subclasses* will become represented more often than in a random choice. As will be shown next, one way of achieving this is by choosing the first partner according to *lowest similarity* to its own class (in  $H_0^{(n)}$ ).

### 3.3.5.4. Minimum similarity choice

Let us denote by  $I_{\text{mins}}$  the version of procedure I, in which the first crossing partner is chosen on the basis of *minimal* occurring

*similarity* of a ramet to its own class (according to  $H_0^{(n)}$ ).

For shortness and clarity, some further notations are first introduced. Whenever needed, the individual ramets in class  $e_u$  are denoted by  $r_u(j)$ ,  $j = 1, \dots, n_u$ , and the ramets in its subclass  $e_{ui}$  by  $r_{ui}(j)$ ,  $j = 1, \dots, n_{ui}$ . Respectively, the ramets in a true class  $e_t$  are denoted by  $r_t(j)$ ,  $j = 1, \dots, n_t$ , and the ramets in the population by  $r(j)$ ,  $j = 1, \dots, n$ . In order to keep the notations simple, let us in what follows consider a certain (though in principle arbitrary) ramet, say  $r_{11}(1)$ .

Now the problem is, what is the *probability* of a *small similarity value* of a ramet to its 'own' class (own as  $H_0^{(n)}$  states), *on condition* that the *classification* represented in  $H_0^{(n)}$  has occurred.

Since cross results are random variables, we could obtain different classifications, when repeating our crossing study (consisting presently of  $n$  crossing cycles) many times. In spite of different actual seed set results, some of the repetitions would yield the same estimated classification ( $H_0^{(n)}$ ) as the present study has given. It is the frequency of a low similarity amongst these "successful" repetitions (yielding a common classification) that the conditional probability under consideration refers to.

This probability can be represented in the form

$$(63) P \{ \text{ramet } r_{11}(1) \text{ exhibits a "low similarity"} \\ \text{to its 'own' class } e_1 \mid \text{classification } H_0^{(n)} \\ \text{has occurred} \} \\ = P \{ S^{(n)}(r_{11}(1), e_1) < z \mid \text{classification } H_0^{(n)} \} \\ = \frac{P \{ S^{(n)}(r_{11}(1), e_1) < z \text{ and classification } H_0^{(n)} \}}{P \{ \text{classification } H_0^{(n)} \}}.$$

Making simplifications in two steps (expressions 63' and 63''), the *behaviour* of this probability can be studied (App. K). The following (approximate) result was obtained.

### Theorem III.

If we *increase* the relative size of a subclass, this will *decrease* the conditional *probability* of any ramet in the subclass to yield a *small (positive) similarity* to its own class (in  $H_0^{(n)}$ ).

In other words, a low similarity to 'own' class is *less common* amongst the ramets of a *greater subclass*. Theorem III holds *at least* for a sufficiently *large* subclass, containing

most ramets of its 'own' class. Very plausibly it will also hold for somewhat smaller subclasses.

We can now study the *effect of choosing* the first partner according to  $I_{\text{mins}}$ . Let us define that case  $A \equiv$  'the similarity value of the ramet to its own class (in  $H_o^{(n)}$ ) is small (though positive)', case  $B_{vb} \equiv$  'the ramet belongs to a (relatively) very big subclass in its class', and case  $B_{mb} \equiv$  'the ramet belongs to a (relatively) moderately big subclass in its class'. Using these notations, Theorem III actually states that  $P\{A | B_{vb}\} < P\{A | B_{mb}\}$ . Utilizing directly the definition of conditional probability, and performing minor rearrangements, it is easy to prove that

$$(72) \quad \frac{P\{B_{mb} | A\}}{P\{B_{vb} | A\}} / \frac{P\{B_{mb}\}}{P\{B_{vb}\}} = \frac{P\{A | B_{mb}\}}{P\{A | B_{vb}\}} > 1.$$

Hence, applying the *condition A*, we can hit at the smaller subclass at a *greater probability* than without applying it.

In short, (even the weak form of) Theorem III implies that using procedure  $I_{\text{mins}}$ , *will* actually *increase* the probability of *hitting at a false relation* (see expression 62). Since, choosing the first partner on the basis of minimal occurring similarity of a ramet to its own class, we will *less often* than using a random choice  $I_r$ , acquire a ramet belonging to any of the relatively largest subclasses, ie. subclasses "dominating" their classes.

### 3.3.5.5. Maximum lability choice

In chapter 3.3.4.3, expression (41''), it was shown that the effect of any single ramet move on the support of a hypothesis can be expressed as a difference between two similarities. Considering  $H_o^{(n)}$ , we can define its *strongest rival*, denoting it by  $H_r^{(n)}$ , to be the hypothesis at a *one ramet move distance* from it, which fulfills the condition

$$(73) \quad \min_j \log T_{oj}^{(n)} = \log T_{or}^{(n)}.$$

Regarding the single ramet involved in the move, say  $r_o$ , the difference

$$(74) \quad \log T_{or}^{(n)} \equiv S^{(n)}(r_o, e_o) - S^{(n)}(r_o, e_r)$$

of its 'similarities' to each of its two competing equivalence classes, will thence be minimal.

We shall define still another term. A ramet  $r$  is said to be the more *stabile* regarding the two equivalence classes  $e_i$  and  $e_j$ , the *greater* will be

$$(75) \quad |S^{(n)}(r, e_i) - S^{(n)}(r, e_j)|.$$

The reverse of 'stability' is called 'lability'. Hence, a ramet is said to be the more *labile* with regard to the classes concerned, when the value of the expression (75) becomes *smaller*. Using such a terminology, the locally optimal hypothesis  $H_o^{(n)}$  and its strongest rival  $H_r^{(n)}$  will differ only via the most labile ramet in  $H_o^{(n)}$ .

•

Choosing the first partner on the basis of *maximal lability*, ie. using the version  $I_{\text{maxl}}$  of the procedure  $I$ , is nearly related to choosing on the basis of minimum similarity to its own class. If the similarity values of the maximally labile ramet to all other classes but its own, happen to be non-positive, then the procedure  $I_{\text{maxl}}$  will be *identical* to the procedure  $I_{\text{mins}}$ . A minimal though positive similarity to own class, will *imply* a high lability, since to the most rivalling class, a ramet will always express at least a zero similarity (which is the similarity of an "emerging" class; see above). In addition to minimum similarity ramets, the set of most labile ramets will also contain ramets, which are positively, even highly (however *almost equally highly*) similar to (at least) two classes in  $H_o^{(n)}$ . Let  $r_1$  be such a ramet; suppose that it has been classified into the class  $e_1$  and expresses a positive similarity also to the class  $e_2$ . This ubiquity of a positive similarity, should serve as a further indi-

cation of a poor classification. Namely, then  $r_1$ -equivalent ramets will usually be fairly frequent also in the class  $e_2$  (see Corollary I).

Concerning maximum lability choice, no investigation was made even as systematic as the one presented in connection with the minimum similarity choice (App. K). For instance, no proof has been given that under procedure I, inclusion of these most ubiquitous cases would further increase the probability of hitting at a false relation (see later). Though in practice, at least during the later stages of the study, the difference between procedures  $I_{\min}$  and  $I_{\max}$  may not be great. However, it is possible that minimum similarity choice might have managed these extra cases more efficiently.

In spite of that uncertainty, the application of *maximum lability choice*  $I_{\max}$ , for finding the *first crossing partner* in forthcoming crosses was used in the present study.

In the background, there was functioning an intuitive, "empiristic" reasoning that attacking at most labile ramets, would be the most evident way to try to change the present optimal hypothesis  $H_0^{(n)}$  into another. Since, even a slight change in the results of such a labile ramet, might be sufficient to remove it from its present class into a rivalling class. Or if instead the similarity of the ubiquitous ramet to its "own" class would increase, then this ramet should at least no longer be included in the set of most labile ramets, in the next crossing cycle.

Another, perhaps a more relevant argument

$$(77) \quad P \{ \text{ramet } r_{11}(1) \text{ is "highly labile"} \mid \text{classification } H_0^{(n)} \} \\ \equiv \frac{P \{ \max_{v=2, \dots, m_0+1} S^{(n)}(r_{11}(1), e_v) - S^{(n)}(r_{11}(1), e_1) > -z, \text{ and classification } H_0^{(n)} \}}{P \{ \text{classification } H_0^{(n)} \}}$$

After some steps, this reduces to

$$(77') \quad P \{ \text{ramet } r_{11}(1) \text{ is "highly labile"} \mid \text{classification } H_0^{(n)} \} \\ \equiv \frac{P \{ 0 \leq S^{(n)}(r_{11}(1), e_1) - \max_{v=2, \dots, m_0+1} S^{(n)}(r_{11}(1), e_v) < z \}}{P \{ 0 \leq S^{(n)}(r_{11}(1), e_1) - \max_{v=2, \dots, m_0+1} S^{(n)}(r_{11}(1), e_v) \}}$$

or, expressed using *standardized similarities*,

$$(77'') \quad P \{ \text{ramet } r_{11}(1) \text{ is "highly labile"} \mid \text{classification } H_0^{(n)} \} \\ \equiv \frac{P \{ \max_{v=2, \dots, m_0+1} S_{\text{std}}^{(n)}(r_{11}(1), e_v) > -z \mid \text{cl. } H_0^{(n)} \}}{P \{ \max_{v=2, \dots, m_0+1} S_{\text{std}}^{(n)}(r_{11}(1), e_v) \leq 0 \}} \\ \equiv \frac{P \{ -z < \max_{v=2, \dots, m_0+1} S_{\text{std}}^{(n)}(r_{11}(1), e_v) \leq 0 \}}{P \{ \max_{v=2, \dots, m_0+1} S_{\text{std}}^{(n)}(r_{11}(1), e_v) \leq 0 \}}$$

was that by utilizing 'stability', it was easy to attain at *standardized similarity* values. Namely, the similarity of a ramet to a given class was standardized by subtracting from it the maximal similarity value of this ramet to any of the classes, ie. subtracting the similarity value of this ramet to its own class, that is

$$(76) \quad S_{\text{std}}^{(n)}(r_{11}(1), e_u) \\ \equiv S^{(n)}(r_{11}(1), e_u) - S^{(n)}(r_{11}(1), e_1) \\ = S^{(n)}(r_{11}(1), e_u) - \max_{v=1, \dots, m_0} S^{(n)}(r_{11}(1), e_v).$$

Hence, the standardized similarity of a ramet to its own class will always be 0, and its similarities to the other classes will be negative (or at most 0). This standardization will render the similarity values of different ramets in different crossing cycles, or in different populations, more readily *comparable*. — Actually, omitting the sign, 'standardized similarity' of a ramet to any equivalence class in  $H_0^{(n)}$ , will be identical to the 'stability' of the ramet in regard to its "own" class and the equivalence class concerned.

Proceeding step by step the respective considerations as those given for minimum similarity choice (3.3.5.4 and App. K), this time in regard with *maximum lability choice*, we will attain at the formulae

By simplifying the expression (77') (in a similar way as in App. K) replacing the  $\max S^{(n)}(r_{11}(l), e_c)$  term by zero, we would once again arrive at the expression (63''). Thence, also Theorem III would apply. Such a simplification would, however, go too far, since this time we would get rid of just those extra cases that we are interested in. Without having these "ubiquitous" ramets (see above) — which in essence is the mere matter rendering procedure  $I_{\max}$  different from  $I_{\min}$  — we could certainly not any more study their effect.

Hence, there remain two possible ways. Either to search for the probability distribution of a maximum of several  $S_{\text{std}}$  values (see 77''); and applying it, work through a proof concerning the behaviour of the probability. Or, to *bypass* such considerations here — regarding them as too complicated; and believe on a merely intuitive basis that the method will function also when a certain number of ubiquitous ramets would be included, ie. ramets "highly labile" however "highly similar" to their own class.

The latter way was chosen, since the aim of the study was to find a practical method of achieving a result with as small amount of labour as possible. Its success or failure can and will be *measured* in part *in practice* — if the optimization is not functioning with full efficiency, then we will be forced to do some extra labour before achieving a final classification.

### 3.3.5.6. Choice of the second partner

Formula (62) above presents the probability of hitting at a false relation, if the second partner is taken *at random* within the equivalence class of the first partner. This probability may be increased, by further utilizing the similarity values. If the second partner is chosen on the basis of the opposite of high lability, ie. on the basis of *non-low stability*, then we will pinpoint ramets, which display a *non-small similarity* to their own class (in  $H_o^{(n)}$ ). Regarding case  $A_c \equiv$  'the similarity value of the ramet to its own class (in  $H_o^{(n)}$ ) is non-small', and referring to the formula (72), it is easy to show that

$$(78) \quad \frac{P\{B_{vb}|A_c\}}{P\{B_{mb}|A_c\}} \cdot \frac{P\{B_{vb}\}}{P\{B_{mb}\}} = \frac{P\{A_c|B_{vb}\}}{P\{A_c|B_{mb}\}} > 1,$$

since Theorem III implies that  $P\{A_c|B_{vb}\} > P\{A_c|B_{mb}\}$ . Hence, applying the condition  $A_c$  in the selection, we can select a larger subclass at a greater probability than without applying it. That is, in comparison to a random choice, utilizing non-low stability in the choice of the second partner, we can *increase* the probability of acquiring a ramet which belongs to a relatively *large* subclass.

Next, in order to study, how this will influence the probability to select a false relation (see expression 62), we will consider the behaviour of the conditional probability  $P\{\text{hit at a false rel.} | 1^{\text{st}} \text{ partner} \in \text{subclass } e_{ui}\}$ . — If  $e_{ui}$  is a *relatively small* subclass, then the second partner will now be less probably than via a random choice, selected from it. Hence, the second partner will be more probably than in a random choice, selected from any other subclass of  $e_u$ . That is, we will *more probably* than before acquire a pair of ramets which according to  $H_o^{(n)}$  should be equivalent, while they actually are not — and thus *hit at a false relation*. — If, on the contrary,  $e_{ui}$  is a *relatively large* subclass, then the second partner will now more probably be selected from it than by using a random choice. Hence, the second partner will be less probably than in a random choice, selected from any other subclass of  $e_u$ . That is, we will *less probably* than before, acquire a pair of ramets which according to  $H_o^{(n)}$  should be equivalent, while they actually are not — and thus hit at a false relation.

The second partner choice concerned, will therefore influence the total probability of hitting at a false relation, in a contradictory way. The actual effect will depend on the actual values of  $P\{1^{\text{st}} \text{ partner} \in e_{ui}\}$  over classes  $u$  and subclasses  $i$ . In the selection of the first partner, procedure  $I_{\max}$  was used to find a ramet which belongs to a relatively small subclass. If procedure  $I_{\max}$  is efficient, then the overall efficiency is further *improved*, when the second partner is selected from the same class on the basis of non-low stability. Without a certain knowledge about the efficiency

of  $I_{\max}$ , however, the question of the profitability of the *non-low stability* choice, remains open.

Actually, the *decision* was made to rely on the efficiency of  $I_{\max}$  in the choice of the first crossing partner, and as a consequence, in the efficiency of the second partner choice concerned. Taking this a step further, and without a direct proof, it was assumed that the efficiency of the selection of the second partner, would be accentuated, if the adjective 'non-low' were replaced by 'high'.

Hence, in the present crossing study, a procedure, which can be denoted by  $I_{\max, \max}$  was utilized, and called a *maximum liability — maximum stability choice*. This procedure thus consisted of choosing the *first* crossing partner on the basis of *maximum liability* between  $H_0^{(n)}$  and its surrounding hypotheses, and the *second* crossing partner from the equivalence class (in  $H_0^{(n)}$ ) of the first partner, on the basis of *maximum stability* between the class and its rivalling classes.

In practice, it was not a pure I type procedure. Since, if the first crossing partner occurred to be chosen from a *uni-ramet equivalence class*, rendering an intraclass choice of the second partner impossible, then the second partner was chosen from its most strongly rivalling class — relying on a type II procedure (see Theorem II).

— As a special case, a ramet which has not yet at all participated in crosses, was formally treated as a maximally labile one. It was preferably crossed to the greatest equivalence class in  $H_0^{(n)}$ .

— In the list of 'Crossing recommendations' (see Fig. 8 and App. E), as a proposed second parent, actually not only the most stabile ramet but also all other ramets in the equivalence class, were printed. This was done for technical reasons, since it could not be known beforehand, which of the ramets in the class would be available in crosses of the next week. This uncertainty was due to great variability occurring in flowering and in the generative quality of flowers, between plots and in the course of the summer. However the ramets in the class were printed out in descending order of stability, each equipped with its liability value (see beneath) and map position in the experimental field.

### 3.3.5.7. Allocation of resources between populations

Considering an arbitrary ramet in class  $u$ , say  $r_u(j)$ , its *standardized similarity* value to its own class will be zero. Therefore, the most

interesting item will be the standardized similarity of the ramet to the *most strongly rivalling class*, say  $er(u, j)$ , that is

$$(79) \quad \text{Lab}^{(n)}(r_u(j)) \equiv S_{\text{std}}^{(n)}(r_u(j), er(u, j)) \\ \equiv \max_{\substack{v=1, \dots, m_0+1 \\ v \neq u}} S_{\text{std}}^{(n)}(r_u(j), e_v).$$

This can be regarded as the *liability of the ramet* between its own class and its most strongly rivalling class. Notice that its values are *negative*, at most zero.

Averaging over the ramets in an equivalence class, we attain at a term, say, *average liability* of ramets, or "need of crossing", in an *equivalence class*  $e_u$

$$(80) \quad \text{Lab}^{(n)}(e_u) \\ \equiv \frac{1}{n_u} \cdot \sum_{j=1}^{n_u} S_{\text{std}}^{(n)}(r_u(j), er(u, j)).$$

Similarly, averaging over all ramets, we attain a value for the "need of crossing in a population"

$$(81) \quad \text{Lab}^{(n)} \equiv \frac{1}{n} \cdot \sum_{\substack{u=1, \dots, m_0 \\ j=1, \dots, n_u}} S_{\text{std}}^{(n)}(r_u(j), er(u, j)).$$

It thus represents the *average liability* of the ramets (between  $H_0^{(n)}$  and the most rivalling class for each ramet) *in the population*. As its other name suggests, this statistic  $\text{Lab}^{(n)}$  can be used to help the allocation of the available crossing resources to the populations under study.

Now, the statistic  $\text{Lab}^{(n)}$  should take care of the allocation of crossing work *between* populations, and the procedure  $I_{\max, \max}$  should take care of the optimized crossing recommendations *within* a population. Though, in order to help in monitoring the advance in classification success during the course of crossing cycles, also another series of *illustrative statistics* was produced. These were printed in the output list 'Equivalence classes of populations estimated' (see Fig. 8 and App. D).

The first of them was the '*dissimilarity of two equivalence classes*'  $e_u$  and  $e_v$ , that is

$$(82) \quad \text{Diss}^{(n)}(e_u, e_v) \equiv -S^{(n)}(e_u, e_v).$$

This was printed out for each pair of equivalence classes in  $H_0^{(n)}$ , in order to see, which equivalence classes were rated to be *least dissimilar* to each other. This statistic was actually utilized in considering, whether two equivalence classes should be amalgamated or not, in our extended single move algorithm. Namely,  $\text{Diss}^{(n)}(e_u, e_v)$  is identical to the criterion function  $\log T_{ij}$  concerning this situation (see expression 45" in 3.3.4.3). Therefore, in  $H_0^{(n)}$ , its value will always be *positive* (if it had not been, then the classes should have been amalgamated).

The second statistic was an *average dissimilarity of an equivalence class* to other equivalence classes,

$$(83) \quad \text{Diss}^{(n)}(e_u) \equiv -\frac{1}{m_0 - 1} \cdot \sum_{\substack{v=1, \dots, m_0 \\ v \neq u}} S^{(n)}(e_u, e_v).$$

This gives a basis to compare, how well "separated" from the other classes, each equivalence class in  $H_0^{(n)}$  actually is.

Finally, averaging over all pairs of equivalence classes, we will attain at the *average dissimilarity of equivalence classes* in the population,

$$(84) \quad \text{Diss}^{(n)} \equiv -\frac{2}{m_0(m_0 - 1)} \cdot \sum_{\substack{u=2, \dots, m_0 \\ v=1, \dots, u-1}} S^{(n)}(e_u, e_v).$$

In addition to  $\text{Lab}^{(n)}$ , this statistic can serve to help to judge, how urgent a crossing within a population is required.

### 3.3.6. Estimation of seed number probabilities

In the partition of ramets into equivalence classes, the quotient function  $p_s(l)/p_c(l)$  was utilized. Since this function is unknown to us, it must be *estimated*. For partitional purposes, the quotient function would suffice. We are, however, interested also in the probability density functions themselves, ie. in both  $p_s(l)$  and  $p_c(l)$ . Hence, both probability density functions are estimated separately.

If one is contented with a parametric estimation, ie. searching the estimated function among a certain, predestined family of curves, such an estimation is usually a straightforward task. The pollination and fertilization in a population is, however, a highly complicated problem (see 1., 2., and TAMMISOLA 1986). Therefore, this study is not restricted to any family of curves predestined by the inevitably incomplete and too far simplified theoretical considerations. Hence, the *non-parametric* methods of density estimation are used.

As pointed out already by WHITTLE (1958), in any method of estimation, some sort of *model* must, however, be postulated. While ordinary curve-fitting is based on a complete stochastic model, in 'curve-smoothing', such a model is only a *partial* one. The model often takes the form of some regularity condition on the curve that is being estimated.

There is a large array of such methods presented in the literature. Conceptually, the most appealing method, searching a density function on the basis of maximum likelihood of the results, unfortunately is known to end in a 'Dirac's catastrophe'. Namely, it will produce a "comb" function with "endless" high peaks at the places of the actually observed results (see eg. LEONARD 1978). This represents an estimated density function with "extremely little smoothing". For almost any conceivable situation, such a comb will be useless, and more smoothing will be required.

Actually, some statisticians (eg. B.W.Silverman in LEONARD 1978, p.134) consider the choice of smoothing method to be less critical than the choice of *how much to smooth*. There are a couple of methods, where the data itself is being used to suggest the optimal rate of smoothing (eg. WAHBA 1977, and an extension of LEONARD 1978, see p.143). That is, the value of a smoothing parameter, which is optimal in the sense that it minimizes the integrated mean square error, can be estimated from the data.

The method of WAHBA (1977) has however been designed only for periodic functions, which renders it unsuitable in this study. Namely, the behaviour of each probability density function in their tail areas will likely be important for the efficiency of discrimination — a periodicity artefact might ruin it. Unfortunately, neither documents nor computer programs of the extended version of LEONARD (1978), were available. If such an extended version really exists, presumably it will at least be more complicated than the published version, and thus difficult to program or time-wasting to use in a computer. Thence, these sophisticated methods with optimized smoothing rate, will be bypassed.

Since we are dealing with an argument (seed number), which accepts only whole number values, we are actually working with discrete probability distributions, not with probability density functions at all. These discrete distributions could in a natural way be estimated via *histograms* representing the real results in crosses. There is however the difficulty that none of the seed number classes is allowed to accept a zero value. Since, this would result in an undetermined or infinitely large value of the quotient. That is, one such case would ruin the classification completely. On the other hand, with moderate numbers of crosses made, such zero classes would be quite common, especially in the tail areas. The relative accuracy of the probability estimates concerning classes with smallest expectations, would also be poor. Therefore — having decided not to restrict ourselves to any *a priori* information in the form of a predestined family of curves — we must rely in part also on information concerning the *nearby* regions, ie. execute some kind of smoothing.

Histogram smoothing methods are available, but unfortunately many of them are known to produce also *negative estimates* (eg. BONEVA et al. 1971) — such a property renders a smoothing method inapplicable to this study. In fact, it was decided to perform an estimation of a continuous density function instead. Due to its simplicity, a modified version of the classical method of PARZEN (1962), was used (see later). The most essential reasons for choosing this way, ie. applying methods purposed for *continuous* random variables, will be presented below.



### 3.3.6.1. Relative seed numbers

The primary random variable is the *number of seeds* produced in a cross, denoted in the present paper either by  $\tau$  or by  $l$ . Two such random variables have been introduced, one concerning an incompatible pollination ('self type' cross, denoted by a subscript  $s$ ) and the other concerning a compatible pollination (denoted by a subscript  $c$ ). Actually the situation is not so simple. Namely, *each population* should require its *own* pair of random variables. Since, the populations may differ in the general fertility of their flowers, in the amount and efficiency of pollen etc. Regarding carefully controlled crosses made with plenty of pollen, the most important difference which remains, is the *different number of pistils* occurring on the flowers of different populations.

Estimating separately the probability density functions for each population, would however have required too much crossing — or the functions would have remained too uncertain and exhibited too much random variation. Therefore, a *modified* random variable, called *relative seed number*, and denoted by  $\tau_r$ , was introduced:

$$(85) \quad \tau_{rs} \equiv \frac{\tau_s}{E\{\tau_s\}}, \text{ and } \tau_{rc} \equiv \frac{\tau_c}{E\{\tau_c\}}.$$

That is, the relative seed number is the actual seed number divided by the average seed number in *compatible* crosses in the population.

In the probability model, it is *assumed* that the *relative* seed numbers are *identically distributed* in different arctic bramble populations. Or that in practice, they are distributed similarly enough that regarding the aims for classification, a *pooled estimate* of  $\tau_{rs}$  and respectively of  $\tau_{rc}$  over populations can be acquired.

When using relative seed numbers, each population will still produce the same number of different cross results as before, but the

relative seed numbers will as a rule *not* coincide for different populations, as the non-modified seed numbers did. Therefore — considering the cross data pooled over populations — while the distribution of relative seed numbers is still in principle a *discrete* one, it will now have a *high number* of different possible values. Having so many classes, it is possible — without severe error — to handle the data as if our random variable would have a *continuous* distribution.

### 3.3.6.2. A modified Parzen estimation method

For independent, identically distributed observations  $X_1, X_2, \dots, X_n$  from a common probability density function  $f$ , PARZEN (1962) derived a class of estimates of  $f$ , called also *kernel* estimates

$$(86) \quad \hat{f}(x) = \frac{1}{nh} \cdot \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right),$$

where  $K()$  is a "hill" function integrating to one and satisfying some regularity conditions, and  $h$  is the parameter determining the *degree of smoothing*. The value of  $h$  remains to be chosen by the examiner usually on the basis of the required degree of visual smoothness in the resulting estimate of the density function.

The 'kernel' method may be more easily understood, if viewed in another way. The "elementary probability mass"  $1/n$  represented by a single observation (out of a total of  $n$  observations), will be "scattered" around the actually realized (observed) value of the random variable. The relative frequencies of this scatter at different distances on either side will be determined by the form of the "hill" or 'weighing function'  $K$ .

In the present study, it was decided to choose  $K$  so that  $(1/h) \cdot K((X_i - x)/h)$  represents the density function of a normal distribution  $N(X_i, h)$ ; ie. the "elementary probability mass" is scattered on both sides

of the observed value obeying a *normal distribution* with its mean at the observed value and with a standard deviation of  $h$  in value. This is achieved by choosing

$$(87) \quad K(y) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2} \cdot y^2}.$$

In this case, concerning relative seed numbers, *negative* argument values will *not occur* at all, while especially with  $\tau_{rs}$ , argument values at or near a zero will be quite frequent. This may cause some trouble, when applying the prescribed hill function. Namely, observations representing very low relative seed numbers, will be “present” in the truly defined area of the density function by about a half of their usual share — almost one half of their “elementary probability mass” will be scattered on a “wastage” region, ie. on negative seed numbers.

Therefore, it was decided to modify the procedure a little. Instead of the ordinary normal distribution, a *truncated* form of it was applied. The normal distribution of scatter at any particular observation was *truncated at zero*, and its function values were increased to compensate the lost area, ie. to retain it as a real density function with an area of one. Actually, a *truncated* normal distribution was thus used: the normal distribution on condition that only non-negative seed numbers will occur. Technically this was achieved by dividing the function values by  $1 - \text{Tail}$ , where Tail denotes the area of the tail at the “wastage” region.

### 3.3.6.3. Return to a discrete distribution

Let us denote the probability *density function* of a *relative seed number* by  $f_r$ . Referring to expression (85), it is straightforward to show that in a *certain population*,

$$(88) \quad \begin{cases} P\{\tau_s = i\} = \frac{(i+1)/E\{\tau_c\}}{i/E\{\tau_c\}} \int f_{rs}(z) dz, \text{ and} \\ P\{\tau_c = i\} = \frac{(i+1)/E\{\tau_c\}}{i/E\{\tau_c\}} \int f_{rc}(z) dz. \end{cases}$$

As already defined above,  $E\{\tau_c\}$  denotes the average seed set in compatible crosses valid in the *population* considered.

Utilizing the slightly modified procedure of kernel estimation, described above, a probability *density estimate* concerning  $f_{rs}$ , and another for  $f_{rc}$ , will be acquired. These density estimates will be defined for all non-negative relative seed numbers, ie. for  $\tau_{rs} \geq 0$  and, respectively, for  $\tau_{rc} \geq 0$ . In this defined region, the function *value* of these density estimates will always be *positive*. Thus they will be suitable for our use in classification purposes, since the value of the quotient function  $p_s(l)/p_c(l)$ , will now always be well defined.

Expression (88) can be used in *estimating seed number probabilities*  $p_s(l)$  and  $p_c(l)$  — which are required in the partition of ramets into equivalence classes — by replacing the unknown quantities with their estimates

$$(88') \quad \begin{cases} \hat{p}_s(l) = \frac{(l+1)/\hat{E}\{\tau_c\}}{l/\hat{E}\{\tau_c\}} \int \hat{f}_{rs}(z) dz, \text{ and} \\ \hat{p}_c(l) = \frac{(l+1)/\hat{E}\{\tau_c\}}{l/\hat{E}\{\tau_c\}} \int \hat{f}_{rc}(z) dz. \end{cases}$$

The integration was made numerically on a B7800 computer, utilizing the quadrature method of PATTERSON (1968 a, b). For this purpose, the program D01ACF in the NAG program library, was extended to proceed at double precision up to 127 points in Gaussian quadrature formulae.

### 3.3.6.4. Estimation of average seed number $E\{\tau_c\}$ in compatible crosses in a population

Expression (88') shows, how a *common density estimate* for  $f_{rs}$  and respectively for  $f_{rc}$ , acquired from the pooled data, can be utilized in each population to yield *population specific* estimates of seed number probabilities. A necessary prerequisite for this is, however, that we have estimated the items  $E\{\tau_c\}$  population specifically, ie. separately in each population concerned.

In most populations, a required number of compatible crosses were available for such an estimation. Namely, in addition to analytical crosses within the population, flowers of the population were also used as seed parents in *control crosses*, thus controlling the quality of pollen in analytical crosses in the other populations. For practical purposes, these control crosses represented compatible crosses, since the sources of pollen were from another populations which were very distant in origin from the population concerned.

In the final, *iterative adjusting* of estimates (see 3.3.7.), also *analytical crosses* were utilized. That is, in the estimation of  $E[\tau_c]$ , the crosses classified as compatible ones according to  $H_0^{(n)}$ , were included as well. In populations containing only a single equivalence class, such intrapopulation compatible crosses are, however, missing.

In only four cases (three vegetatively or generatively weak populations, 063A, 089B and 204D, and a uniclinal population 002A), the total number of compatible crosses remained too low to form a solid basis for the estimation in question. In these cases, a more reliable estimation of  $E[\tau_c]$  was acquired via linear regression.

Since, seed numbers data from *open-pollinated*, non-emasculated flowers, were more readily available in many populations. When recording the fruit set in the plots several times through the summer, also seed set recordings were made on each plot. The number of seeds in (at most) five greatest berries per plot, were counted each time the berries were harvested. Into these records, a regression line  $y = \beta \cdot x$  was fitted. Here  $y$  denotes the average seed number in compatible crosses in the population,  $\beta$  is the regression coefficient, and  $x$  is the average number of seeds in the five greatest berries per plot in the population at each time. In calculating the estimated value of the regression coefficient, data from different populations was weighed by the square root of the number of crosses concerned in each.

The estimated value of the regression changed only slightly between different cycles of iteration. A typical value occurring was  $\beta \approx 0.73^*$ .

### 3.3.7. Iterative adjusting of equivalence class and seed number probability estimates

During the *crossing phase* in the experimental field, up to and including summer 1981, seed number probability distributions remained unchanged. They were based on cross

results acquired before summer 1978, on which a graphical smoothing was performed. These *preliminary* seed probability *distributions* are presented in the Tab. 6 and in Fig. 17 (see below).

When the field work phase was completed in 1981, the analysis portion of the computer system (see Fig. 8) was modified to cope with changing and *population specific* probability distributions. Then an *iteration phase* was started (Fig. 12). Utilizing the current seed number probability estimates, the ramets in each population were partitioned once again into equivalence classes. These updated equivalence class estimates in each population, were then used — in the manner described above (see 3.3.6.) — for acquiring new seed number probability estimates. This cycle was then repeated until a *convergence* was achieved, that is, until no change occurred in the equivalence class estimates of any population. — No theoretical guarantee could be given for a convergence to occur. In different trials, however, a convergence was always achieved (see 4.1.5.).

Since the present method of clustering — following the terminology of GOOD (1977) — is of the type *not* exercising an *exhaustive search*, there remains the possibility that we succeed in finding only a *local*, not a *global*

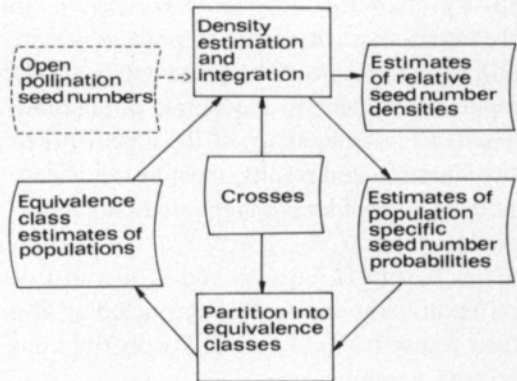


Fig. 12. Iteration cycle to adjust the equivalence class and seed number probability estimates. Crossing phase has already been completed, and the final information concerning the crosses, is obtained from a disc in the computer.

maximum point of the clustering criterion function. Therefore, in the present study, altogether *four different* iterations, beginning at widely different starting points, were performed.

### 3.3.7.1. Choice between end results of different iterations

The actual analysis programs before the onset of the iteration phase, did *not* calculate the actual *likelihoods* of different hypotheses, but only their likelihood ratios. These are of course computationally much simpler, and will suffice for the partitioning made by the single move algorithm. However, even though the algorithm was a slightly "extended" one, such likelihood ratios were commonly acquired only for pairs of "adjacent" hypotheses, at a distance of no more than a one ramet move from each other.

In order to make a comparison of two iteration results, a new computer program had to be coded. Such a program must deal with *arbitrary partitions*, ie. with cases where the two partitional hypotheses may be separated by a distance of more than a single ramet move. Another important difference to be coped with, is that the *seed number probabilities* will usually *not* be *identical* in both hypotheses, since during each iteration, they are *estimated* as well. In addition, in both hypotheses, these probabilities are *population specific*. In the iteration, crosses from all populations are utilized in the estimation of seed number probabilities pertinent to any single population. Therefore, a comparison of the superiority of two iteration end results, must be made considering *all populations* in the study *simultaneously*.

Let  $H_i$  and  $H_j$  be such end results of two different iterations — this time including also their respective final probability distributions of seed numbers.

The ratio of total likelihoods in the 'general' method would be

$$(89) \quad \frac{L_i}{L_j} = \frac{C_i}{C_j} \cdot \sum_k \left\{ \frac{\prod_l p_{iks}(l)^{m_{iks}(l)} \cdot \prod_l p_{ikc}(l)^{m_{ikc}(l)}}{\prod_l p_{jks}(l)^{m_{jks}(l)} \cdot \prod_l p_{jkc}(l)^{m_{jkc}(l)}} \right\},$$

where the index  $k$  runs over all populations in the study.

Making a logarithmic transformation and neglecting the coefficient part, we will finally obtain the utilized, 'short hand' criterion function

$$(90) \quad \log T_{ij} = \sum_k \log T_{ijk},$$

where in general

$$\begin{aligned} \log T_{ijk} &\equiv \sum_l m_{iks}(l) \cdot \log p_{iks}(l) \\ &- \sum_l m_{jks}(l) \cdot \log p_{jks}(l) \\ &+ \sum_l m_{ikc}(l) \cdot \log p_{ikc}(l) \\ &- \sum_l m_{jkc}(l) \cdot \log p_{jkc}(l), \end{aligned}$$

or, for any population  $k$ , which is partitioned in an *identical* way in both hypotheses  $H_i$  and  $H_j$ ,

$$\begin{aligned} \log T_{ijk} &\equiv \sum_l m_{iks}(l) \cdot [\log p_{iks}(l) - \log p_{jks}(l)] \\ &+ \sum_l m_{ikc}(l) \cdot [\log p_{ikc}(l) - \log p_{jkc}(l)]. \end{aligned}$$

As described before (see 3.3.4.), the hypothesis  $H_i$  will be supported over  $H_j$ , if and only if  $\log T_{ij} > 0$ . Hence, considering these two hypotheses, on condition  $\log T_{ij} > 0$ ,  $H_i$  will be chosen. That is, out of these two hypotheses, it will then constitute the *least local uncertainty* (LLU) estimate (see 3.3.3.) of a partition into equivalence classes as well as of seed number probabilities.

## 4. Results

### 4.1. Grouping of the sampled ramets into incompatibility classes

#### 4.1.1. Crosses

During the summers 1978—1980, in total 4027 crosses were made. Almost one half of these were control crosses, which were made in order to ascertain the quality of the pollen flower, i.e. the general fertilizing ability of its pollen. While the 'analytical' crosses were always made within a population, their control crosses were as a rule made with flowers of other populations.

The routine crossing work was a very slow procedure. In favourable weather conditions in June — including the time expended in control crosses, note keeping, cleaning out, and planning of crosses — the most proficient students could achieve on the average 4—5 crosses per hour.

Controlling the quality of pollen proved essential. Since, particularly later in the summer, i.e. from July to September, the fertilizing ability of the pollen from different flowers varied remarkably. On the basis of the damage on the flower caused by eg. thrips, powdery mildew or rain, the low pollen quality could sometimes be predicted. Quite often these low quality pollen flowers (and occasional deficiencies in a crossing technique as well) could only be revealed due to the control crosses.

A minor proportion of the crosses were unsuccessful due to other factors. For instance, the label was lost, the isolation bag had disappeared (often due to young birds tearing them) or opened, the shoot had been broken (due to a careless student or often a weevil), the pistils had been damaged during emasculation, too young buds had been chosen, or there were mistakes in the crossing records. Depending on the persons and their experience, such errors by an isolating or crossing student occurred at a frequency from over five per cent (by a careless novice) to less than one per cent (by a fully trained, careful expert).

Discounting control crosses and crosses which failed due to the above-mentioned reasons, the net number of analytical crosses in this study amounted to 1208 crosses. On the basis of their results, the sampled ramets were then grouped into equivalence classes of incompatibility. Considering the 21 most thoroughly analyzed populations, on average 56 net crosses were made per population (Tab. 5).

#### 4.1.2. No self-fertile clones were found

Among the analytical crosses in the populations — and also as a separate procedure with the correspondents' accessions — about two hundred true self-pollinations were performed. Their results were in part utilized in the iteration stage, for the initial distribution of the seed number probabilities in an incompatible cross (see later). In this sample of accessions studied, *not a single* arctic bramble clone proved self-fertile.

The bulk of the accessions in the experimental field unfortunately remained untested in this respect. Among them a self-fertile clone — desirable in regard to plant breeding purposes — may still have been found.

On the basis of the sample studied, it is statistically *highly unlikely* that in Finland any self-fertile arctic bramble clones would *commonly* exist. This result is contrary to what eg. SALMINEN (1948) and VAARAMA (1965) speculated.

As usual, a "rescue hypothesis" could of course still be offered, stating that in *other* environmental conditions — at their original sites in nature — these tested clones would have revealed self-fertile instead.

If self-fertility were a common phenomenon, also *homozygous* individuals with regard to the incompatibility locus should have been met. Searching for *s* allele homozygotes, occasionally reciprocal crosses were made (see

Table 5. Job statistics in grouping

Population	Net no. of crosses	No. of half-diallel crosses <sup>1</sup>	No. of single moves	No. of amalgamations	No. of classified ramets	Average lability <sup>2</sup>	Max lability <sup>3</sup>	Min lability <sup>4</sup>	No. of equivalence classes	No. of ramets in equivalence classes
002A	33	406	25	4	29	-3.0	-1.3	-22.	1	29
004A	141	435	23	0	30	-3.4	-5.4	-16.	7	12,8,4,3,1,1,1
004B	43	28	4	0	8	-88	-1.3	-5.3	4	3,3,1,1
010A	102	435	25	0	30	-4.9	-1.1	-25.	5	23,4,1,1,1
021A	48	105	10	0	15	-1.6	-11	-13.	5	8,2,2,2,1
030A	20	36	8	0	9	-68	-88	-5.3	2	6,3
032C	64	325	23	0	26	-2.3	-13	-19.	4	19,4,2,1
037A	11	10	3	0	5	-49	-1.3	-2.6	2	3,2
038B	35	351	25	1	27	-3.2	-1.1	-29.	1	27
039A	64	378	27	0	28	-3.2	-1.2	-15.	2	25,3
042A	64	210	16	0	21	-3.2	-1.1	-9.9	5	6,6,6,2,1
075A	38	210	17	1	21	-1.9	-48	-9.6	3	13,5,3
078A	111	253	16	1	23	-3.3	-66	-18.	6	13,5,2,1,1,1
087A	19	45	7	0	10	-1.4	-1.1	-7.4	3	8,1,1
089B	31	378	25	2	28	-2.6	-56	-18.	1	28
099A	40	378	27	2	28	-3.3	-1.2	-21.	1	28
102E	55	231	16	1	22	-1.4	-54	-5.7	5	11,5,2,2,2
115B	32	406	23	5	29	-2.9	-1.3	-12.	1	29
116A	—	435	—	—	30	—	—	—	1	30
186A	137	378	24	0	28	-4.7	-88	-9.5	4	9,8,7,4
190A	37	435	28	3	30	-3.2	-1.2	-21.	1	30
212A	42	378	23	2	28	-2.3	-99	-17.	3	22,3,3
Total	1167	6246	395	22	505				67	
Average	56	284	19	1	23	-2.6 <sup>a</sup>	-.90 <sup>a</sup>	-14. <sup>a</sup>	3	

<sup>1</sup> The number of crosses which would be required in partitioning a respective number of individuals using a conventional half-diallel design in a *deterministic* system

<sup>2</sup> Lab<sup>(final)</sup> (see expression 81)

<sup>3</sup> Max Lab<sup>(final)</sup>(r) (see expression 79), considering only classified ramets r

<sup>4</sup> Min Lab<sup>(final)</sup>(r) (see expression 79), considering only classified ramets r

<sup>a</sup> The arithmetic mean of logarithmic values, ie. a logarithm of the geometric mean

3.1.). The clones were also examined for any signs of inbreeding depression. No such homozygotes could be found, however.

#### 4.1.3. Grouping procedure

After the results of the new crosses were recorded and included into the register of cross results, the grouping algorithm was always started anew *from the beginning*. If one had tried to *continue* the classification from the final grouping of the previous computer run, some computer resources could have been saved. However, since a single move method may find a local optimum only, continuing always from the old grouping could have

limited its possibilities to find out the true *globally* optimal subdivision into incompatibility classes. A "cheaper" but (possibly) worse estimate of equivalence classes would yield *less efficient* crossing recommendations. Therefore — in order to save human *crossing labour* as much as possible — some wasting of machine resources was justified. As a subsidiary effect, the logic of the procedures remained simpler. Thence, considering the thousands of lines of code in the programme, a better reliability of the programmes was also achieved.

On average the classification of 23 sampled ramets in a population could be finished. This partition required on average 19 single moves

and one amalgamation of groups (Tab. 5). Hence, the grouping procedure yielded the subdivision almost with a minimal number of moves possible — regarding a start from the beginning. The extension portion, ie. the amalgamation of groups, proved profitable especially in uniclassic populations. In them, without the extension, one would often have been caught in local (but not global) optimum points. Hence, crossing work would have been guided with essentially less efficient recommendations. By amalgamating the solely artificially separate groups, the classification process in uniclassic populations could be intensified apparently in a similar way as using procedure II in them (see App. J, Example 1).

The output of the grouping procedure (App. D) contains the estimates of the equivalence classes and certain statistics for classification, including the standardized similarities of the individual ramets to different equivalence classes (expression 76), the liability values of the individual ramets (expression 79), of equivalence classes (expression 80) or of a population (expression 81), and the pairwise dissimilarities of the equivalence classes (expression 82). Hence, this output list could have been utilized in designing the forthcoming crosses. In practice this was, however, unnecessary, since the most important of these statistics were also printed in the list of crossing recommendations.

#### 4.1.4. Crossing recommendations

New crosses to be made were selected using the list of crossing recommendations (App. E). The crosses proposed in it were chosen and arranged on the basis of the principle of 'maximum liability — maximum stability' (see 3.3.5). The relatively small numbers of net crosses required (Tab. 5) imply that crossing recommendations proved to be fairly efficient in practice. Therefore, though the recommendation algorithm has to be considered semi-heuristic — due to the gaps in the proofs — it still seems to be based on a solid foundation.

In fact, all trials to *speed up* the equivalence classification via human *intuition* proved unsuccessful. For a while, ramets with apparently similar morphology were regarded most likely to belong into a common equivalence class. Hence, choosing such ramets as crossing partners, would result in a quicker classification of ramets into their true equivalence classes. However when this was put into practice, a clear *retardation* in the progress of the classification process resulted. Hence, more favourable results were obtained by applying the crossing recommendations instead. One reason for the inefficiency of such an intuitive approach may have been the inability of the students to recognize, which morphological differences should be considered essential in arctic bramble — this would have required special training with well known clonal material. Another reason may have been that a human person has difficulties to keep track and extract the information contained in a large number of different crosses. For instance in population 186A, where intuitive methods were used for a short trial period, a full partition required somewhat more crosses than expected. Therefore, Fig. 14 gives a somewhat conservative picture of the efficiency of the guidance system.

The numbers of crosses actually required in the present study, were compared to the numbers needed *a priori*, if a respective partition had been made applying an ordinary half-diallel cross scheme (without selfings) in a deterministic system (Tab. 5). Utilizing the guidance system, even about 1/5 or 1/10 of the number required without an optimization, would suffice.

Although self-incompatibility is quite strict in arctic bramble in comparison to many other plant species, the number of seeds in its berries should still be considered a *stochastic* variable (Fig. 17). Since, even a controlled compatible cross will every now and then yield no seeds at all, and an incompatible cross may even sometimes produce several seeds. In part this will result from technical errors, a certain

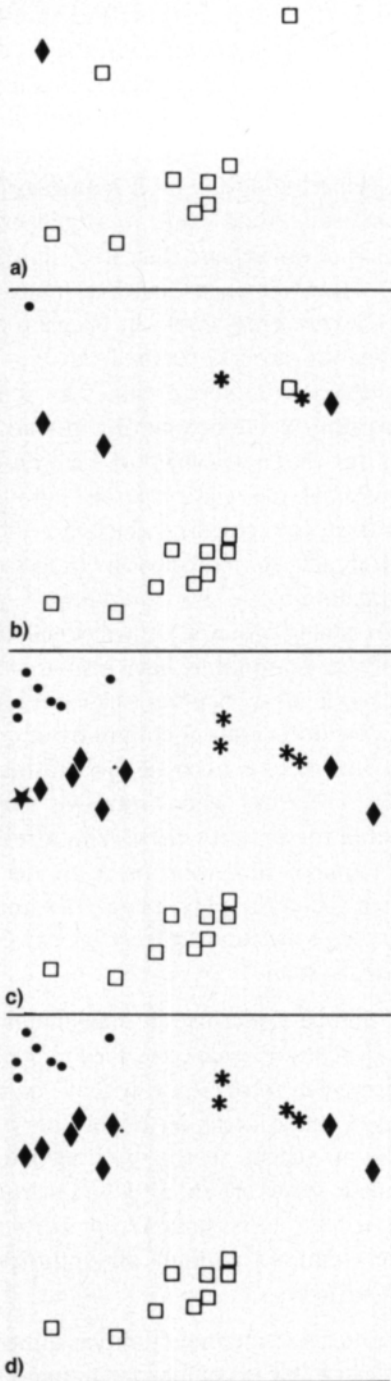


Fig. 14. Progress in analysis of population 186A into equivalence classes. Equivalence class estimates in a map after a) 24 (net) crosses, b) 46 crosses, c) 127 crosses, d) 137 crosses and final iteration. Different symbols denote different equivalence classes (except the single small dot, which represents an individual remaining 'unanalyzed' i.e. non-classifiable due to missing crosses).

proportion of which will inevitably remain undetected by the controls.

Since a classification in a *stochastic* system would require much more crosses than in a deterministic one, still essentially *better gain* than that presented above was actually achieved. Though in a stochastic system, it will be difficult to decide, when a classification should be considered 'ready'. Still it will be safe to state that utilizing the guidance system, the number of crosses required could be diminished to less than  $1/10$  of the number needed without optimization — and in uniclassic populations presumably even lower.

It was planned that the crossing resources should have been allocated between the populations on the basis of the *average lability* of their ramets (expression 81). However this aim was not completely fulfilled, because different populations remained at different levels of average lability (Tab. 5). This was primarily due to technical difficulties. Since, in some populations, or in some individual plots, not enough flowers were available. In addition, some populations of a northern origin matured earlier than expected.

The boom in flowering usually started at the end of April and ceased before the end of June. During this period, instead of minimizing the number of crosses, the crossing recommendations were as a rule applied so as to minimize the total *duration* of crossing study. Accordingly, in the recommended order, as *many* crosses as were technically possible, were always performed.

One natural restriction was posed by the length of the working day. Normally, crossing could be started at about 8.30 — 9 a.m. Before that time, any fresh pollen was generally not yet available. Respectively, most crosses made after 8 — 9 p.m. proved unsuccessful — maybe due to deficiency of fresh pollen or to high humidity at night.

The guidance system for crosses automatically decreased the lability of all the most ambivalently classified individuals. Since as first partners in forthcoming crosses, specifically the most labile individuals were chosen. In many populations, some individual plants remained unclassified (Tab. 5). Therefore, in



comparison to the average lability of all individuals in a population ( $-2.6$ ), the lability of the most ambivalently classified individual remained on average considerably high ( $-0.90$ ). These lability values are however logarithmic ones (expressions 79—81, 76 and 41''—43''). Therefore, even concerning the most lably classified individual, the likelihood of the chosen equivalence class estimate will be on average almost *tenfold* in comparison to the best competing classification (at the one move distance). Regarding other individuals, the likelihood of the estimate was generally much more superior to its rivals — considering the most stable individual, on average even  $10^{14}$ -fold.

Furthermore, one should notice that considering labilities, the 'mean' values in Tab. 5 actually represent *geometric* means. Since the values near zero will in fact become overweighted when calculating these means. Hence, considering *arithmetic* means, the average labilities would be *smaller* than those presented above.

#### 4.1.5. Iterations

After finishing the crossing stage, seed number probability distributions and equivalence class estimates were adjusted iteratively (3.3.7., Fig. 12). In principle, such iterations could have been performed earlier, eg. as a component part of each crossing cycle. In practice this could not be realized in the present study, since the iteration procedures were coded just after having completed the crossing stage.

The probability densities of relative seed numbers were estimated using a modification of the 'kernel' estimation method of Parzen (see 3.3.6.2). A value 0.1 was fixed for the *smoothing parameter*  $h$ . Since in most populations a compatible pollination yields on average about 30 seeds, such a value of  $h$  will be equal to a standard deviation of about three seeds in the "hill" function. With this value of a smoothing parameter, maybe "too well" behaved probability distributions were achieved (Figs 17 and 15). However, the degree of "smoothness" was not considered an especially important issue in this application.

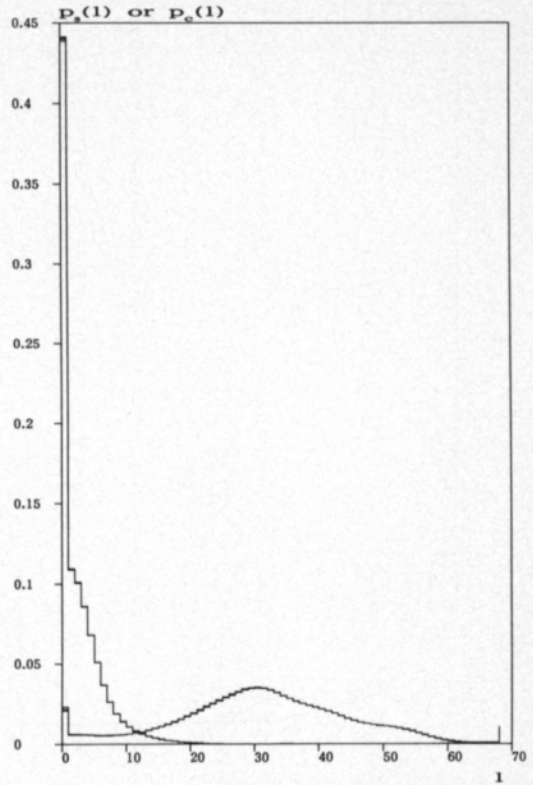


Fig. 15. Seed number probability distributions in population 039A: the convergence results of different iterations (4, 35, 57 and 81). The end results of different iterations appear to be nearly identical and hard to distinguish from each other visually.

In any case, no time was available for thorough experimentation with different  $h$  values. For the numerical integrations of the estimated densities — yielding the seed number probabilities — the quadrature method of Patterson was selected (see 3.3.6.3), principally due to its general applicability and security. Errors in integration could have ruined the equivalence classification and thence the efficient guidance of the crossing work, or at least the time-table of the study. Once again, somewhat more computer resources were utilized in order to increase the certainty. Accordingly, no problems occurred in the integrations.

In order to acquire an adequate view of the convergence of the iteration, altogether four different iterations were performed (Figs 16 and 15). Each of them started from different

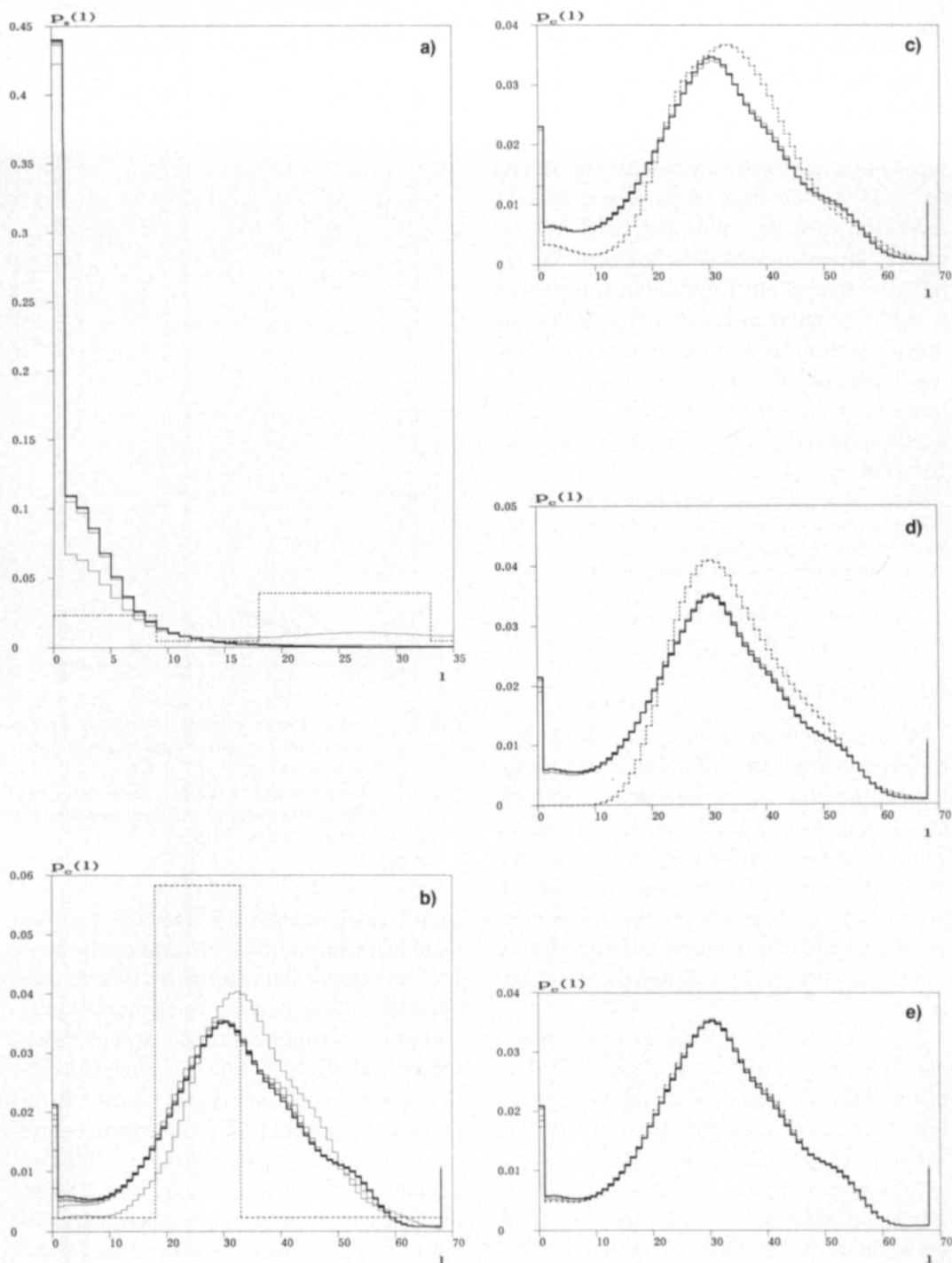


Fig. 16. Change of seed number probability distributions in population 039A during different iterations. a) Equivalent cross, cycles 50—57, b) non-equivalent cross, cycles 50—57, the (arbitrary) initial distribution is denoted by a dashed line, c) non-equivalent cross, cycles 1—4, d) non-equivalent cross, cycles 31—35, and e) non-equivalent cross, cycles 77—81. Thick line: distribution at convergence.

initial conditions, ie. from different initial distributions of the relative seed number probabilities.

Iteration "77" was launched directly on the basis of the end result from the pre-iterative stage. As was the rule in the crossing stage, the equivalence class estimates had been acquired applying the preliminary seed number probability distributions (see later and Figs 17 and 18). By re-interpreting the crosses as intraclass ( $\tau_{is}$ ) or interclass ( $\tau_{ic}$ ) ones (see expression 85) according to these equivalence class estimates, the starting probability distributions were provided. A convergence was achieved after five iteration rounds (77,78,...,81).

In starting the iteration "31", the initial distributions of relative seed number probabilities were acquired in the following way. All controlled crosses in the register were inspected. The crosses yielding at most three seeds, were interpreted as intraclass ones, while the crosses yielding at least 15 seeds were regarded as interclass ones. Other crosses were omitted. A convergence was once more achieved after five iteration rounds (31,32,...,35).

The starting distributions for the iteration "1" were constructed in the following manner. Out of the controlled crosses, all true selfings (ie. crosses within a plot) as well as all crosses in certain non-fruiting populations (002A, 038B, 099A, 115B or 190A) were accepted as 'equivalent' crosses. Only the crosses made between different populations were regarded as 'non-equivalent' ones. Now a convergence was achieved already after four iteration rounds (1,2,3,4).

Aiming at divergence, arbitrary and — with respect to intraclass crosses — even "exotic" step functions (Fig. 16) were tried in the iteration "50". As a surprise, even this iteration converged, though only after eight rounds (50,51,...,57).

Unfortunately, due to a computer change at the Computer Centre, further trials with still more exotic initial distributions could no

more be made. Hence, no instance of a divergence in the iteration was produced.

#### 4.1.6. Estimates of probability distributions

During the *crossing stage*, only *fixed* seed probability distributions (identical for each population) were used. In principle, population specific distributions (see beneath) would have yielded a somewhat more efficient guidance of crosses. However, in practice, the refinement of computer programs required for their estimation, could only be made at a later date.

These preliminary distributions were obtained from "secure" cross results before the year 1978. The frequency histograms were smoothed by graphical means. For most populations, the resulting preliminary probabilities (Tab. 6) did not deviate very much from the final ones obtained in the study (Figs 17 and 18). However for certain populations with exceptional seed number averages (eg. 089B), the discrepancy was more fundamental.

Later on, in the *iteration stage*, the probability model was made somewhat more general. That is, each population was provided with its *own* probability distributions for seed numbers. However these distributions were allowed to differ between populations only with regard to their *mean value*. That is, a common distribution of *relative* seed numbers was postulated (see 3.3.6.1). Such a generalization proved necessary, since considerably different seed number averages were recorded in different populations (Tables 7 and 8).

Each of the four iterations (1, 31, 50 and 77) converged to almost *identical* probability distributions (Figs 17 and 15), which were practically inseparable in a graphic representation. However, in three cases out of the four, the most important quantity in classification, ie. the *ratio* of seed number probabilities in an equivalent versus non-equivalent cross, displayed clear differences (Fig. 18). But

Table 6. Preliminary seed number probability distributions  $p_s(l)$  and  $p_c(l)$  utilized during the crossing phase<sup>1</sup>.

N:o of seeds $l$	Seed number probability in a pollination		$\frac{p_s(l)}{p_c(l)}$	$\text{Log}_{10} \left[ \frac{p_s(l)}{p_c(l)} \right]$
	Intra- class $p_s(l)$	Inter- class $p_c(l)$		
0	0.374	0.0259	14.4	1.16
1	0.205	0.00714	28.7	1.46
2	0.128	0.00551	23.2	1.37
3	0.0692	0.00487	14.2	1.15
4	0.0487	0.00519	9.38	0.972
5	0.0333	0.00551	6.04	0.781
6	0.0205	0.00649	3.16	0.500
7	0.0154	0.00779	1.98	0.296
8	0.0115	0.00908	1.27	0.104
9	0.0103	0.0104	0.988	-0.00530
10	0.00897	0.0120	0.748	-0.126
11	0.00769	0.0139	0.551	-0.258
12	0.00641	0.0156	0.412	-0.385
13	0.00564	0.0175	0.322	-0.492
14	0.00513	0.0195	0.263	-0.579
15	0.00385	0.0214	0.180	-0.746
16	0.00308	0.0234	0.132	-0.880
17	0.00282	0.0250	0.113	-0.947
18	0.00256	0.0272	0.0941	-1.03
19	0.00231	0.0292	0.0790	-1.10
20	0.00205	0.0311	0.0659	-1.18
21	0.00192	0.0331	0.0581	-1.24
22	0.00187	0.0344	0.0544	-1.26
23	0.00180	0.0357	0.0503	-1.30
24	0.00174	0.0363	0.0480	-1.32
25	0.00169	0.0367	0.0462	-1.34
26	0.00164	0.0360	0.0456	-1.34
27	0.00159	0.0357	0.0445	-1.35
28	0.00154	0.0350	0.0439	-1.36
29	0.00149	0.0344	0.0432	-1.36
30	0.00144	0.0337	0.0426	-1.37
31	0.00138	0.0324	0.0427	-1.37
32	0.00133	0.0315	0.0424	-1.37
33	0.00128	0.0302	0.0425	-1.37
34	0.00123	0.0285	0.0431	-1.37
35	0.00118	0.0269	0.0438	-1.36
36	0.00113	0.0250	0.0452	-1.35
37	0.00108	0.0234	0.0461	-1.34
38	0.00103	0.0211	0.0486	-1.31
39	0.000974	0.0185	0.0527	-1.28
40	0.000923	0.0169	0.0547	-1.26
41	0.000872	0.0130	0.0672	-1.17
42	0.000820	0.0110	0.0744	-1.13
43	0.000769	0.00908	0.0847	-1.07
44	0.000641	0.00714	0.0898	-1.05
45	0.000513	0.00519	0.0988	-1.01
46	0.000385	0.00292	0.132	-0.880
47	0.000256	0.00195	0.132	-0.881
48	0.000128	0.00130	0.0987	-1.01
$\geq 49$	0.000385	0.0195	0.0198	-1.70

<sup>1</sup> Based on results of pilot crosses not later than 1978 (graphically smoothed).

Table 7. Average seed numbers of populations in compatible and incompatible pollinations<sup>1</sup>.

Population	$\bar{x}_c$	$\bar{x}_s$	N:o of crosses utilized in estimation	Net n:o of analytical crosses
002A	16.7 <sup>a</sup>	0.03	37	33
004A	24.3	3.10	112	134
004B	24.0	0.00	41	43
010A	24.1	2.50	75	99
021A	20.2	1.00	46	46
030A	16.1	0.63	18	19
032C	21.4	0.48	61	58
037A	28.4	0.00	24	11
038B	30.5	1.07	62	35
039A	30.9	1.00	49	62
042A	23.9	0.79	63	64
051A	22.3	4.00	11	11
063A	19.0 <sup>a</sup>	1.40	5	5
066A	—	—	—	—
066D	—	—	—	—
075A	11.1	1.58	37	37
078A	23.8	1.97	109	111
087A	31.0	1.25	34	19
089B	3.6 <sup>a</sup>	0.10	31	31
090A	22.0	1.00	10	10
099A	22.0	1.42	36	39
102E	19.5	0.83	52	53
115B	24.1	0.03	38	32
116A	—	—	—	—
186A	25.1	1.38	128	132
190A	31.5	0.57	82	37
204D	12.3 <sup>a</sup>	0.00	3	2
212A	25.3	0.82	45	39
304A	—	—	—	—
Pooled	23.6	1.20	1209	1162

<sup>1</sup> Estimated iteratively on the basis of control crosses and analytical crosses.

<sup>a</sup> Estimated on the basis of open pollination seed numbers (see 3.3.6.4 and Tab. 8).

this deviation is restricted to the region of *numerous* seeds. In an equivalent cross, such great numbers of seeds are extremely rare, and therefore their probability estimates ( $\hat{p}_s(l)$ ) are provided with very large *relative* error. Considering the classification process in practice, such differences will only have a negligible importance. Since, crosses producing such great numbers of seed will in any case be powerfully enough classified as non-equivalent ones.

According to the likelihood ratio criterion (see 3.3.7.1 and below), the probability estimates provided by the different iterations — together with their respective equivalence class partitions — were compared. The estimates

with the least local uncertainty were provided by the iteration 50.

#### 4.1.7. Estimates of equivalence classes

All four iterations ended in almost *identical* partitions into equivalence classes. Differences existed only in the classification of one or two ramets in one or two populations. Regarding population 078A, the iterations 1 and 31 produced identical partitions, as did the iterations 50 and 77 with each other. Classification by iteration 50 differed from the classification by iteration 1 in that two small equivalence classes were amalgamated to one; in addition, one extra ramet appeared analyz-

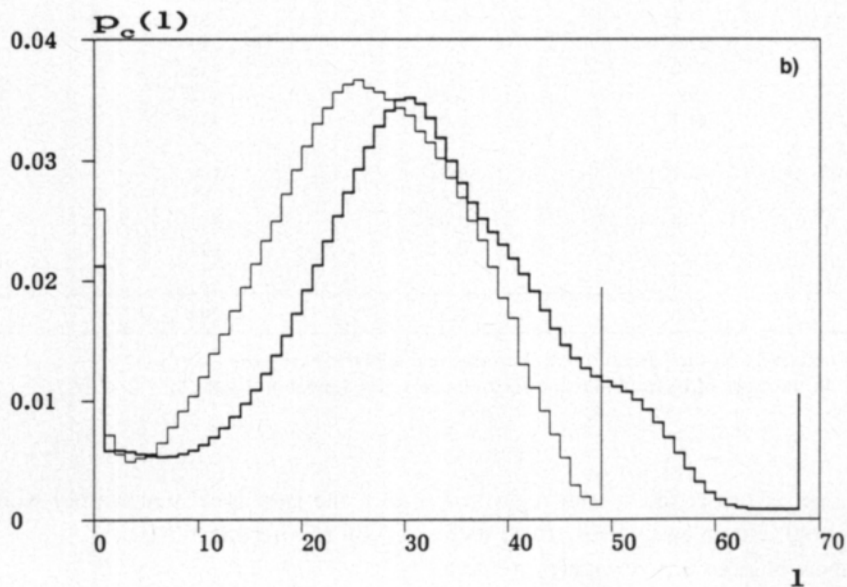
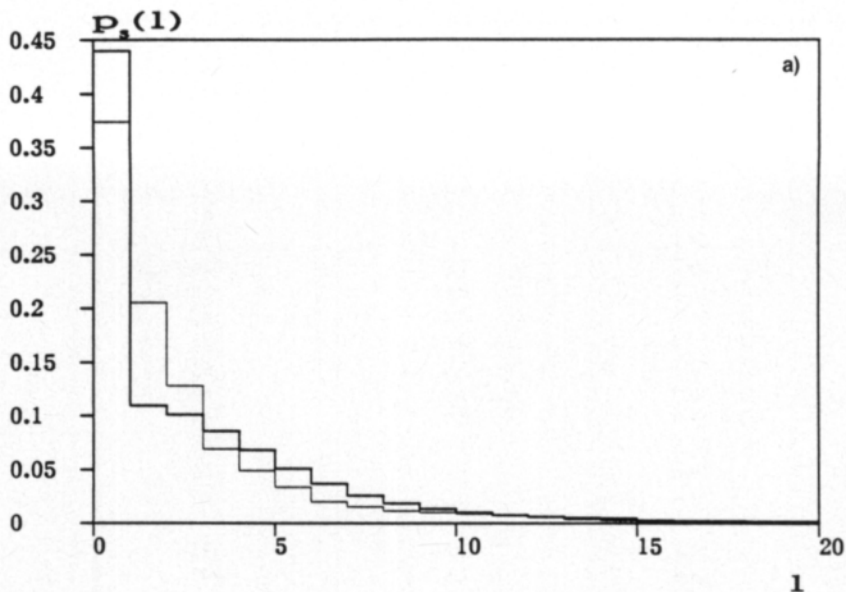


Fig. 17. Seed number probabilities in population 039A.

a) Equivalent cross, ie. a cross within an incompatibility class, and b) non-equivalent cross, ie. a cross between two incompatibility classes. Seed number is denoted by  $l$ .

Fine line: the preliminary distribution utilized in the crossing phase. Thick line: final distribution after iteration.

able, constituting a new equivalence class. In population 186A, the partition by iteration 1 was different from the others, since one ramet from its second greatest equivalence class had been abscised into a new, emerging class.

The resulting partitions from different iter-

ations (together with their respective probability distributions) were compared. The evaluation was based on the likelihood ratio criterion (see 3.3.7.1). In the meaning of least local uncertainty (LLU), the convergence result of iteration 50 (from round 57) yielded

Table 8. Open pollination seed numbers of populations in the experimental field<sup>3</sup>.

Population	$\bar{x}$	S.E.M.	n
002A	22.9	0.5	182
004A	28.7	0.6	153
004B	30.8	0.7	97
010A	34.6	0.7	170
021A	28.4	1.0	65
030A	26.0	0.6	167
032C	33.8	0.8	123
037A	39.4	2.2	28
038B	34.7	0.6	140
039A	45.3	0.9	130
042A	32.6	1.0	88
051A	23.0	1.5	28
063A	26.0	3.3	15
066A	—	—	—
066D	—	—	—
075A	29.6	0.7	94
078A	36.7	0.8	163
087A	31.9	1.0	60
089B	5.0	0.3	57
090A	25.4	1.5	20
099A	23.3	0.6	155
102E	26.4	0.6	212
115B	32.5	1.0	134
116A <sup>1</sup>	1.8	0.2	26
186A	33.8	0.7	145
190A	36.8	1.0	135
204D	16.9	1.7	30
212A	32.7	1.5	51
304A	—	—	—
Pooled	30.5	0.1	2642
Weighted <sup>2</sup>	31.6	0.2	

<sup>1</sup> Based on fruit set observations covering the whole of summer 1980. Not included in total averages.

<sup>2</sup> A weighted mean of population averages, using as a weight for a population its net number of analytical crosses (see Table 7)

<sup>3</sup> Up to five greatest berries per plot are counted. The numbers are based on the first picking in 1980.

the best combined estimate of equivalence class partition and seed number probability distributions in the entire population data in our study. The next best proved to be the result of iteration 77 (from round 81). Their equivalence class estimates were completely identical, but their estimated seed number probability distributions differed a little from each other. Their likelihood ratio (see expression 90) was  $T_{57,81} = 1.94$ . The results of iteration 31 (round 35) and iteration 1 (round 4)

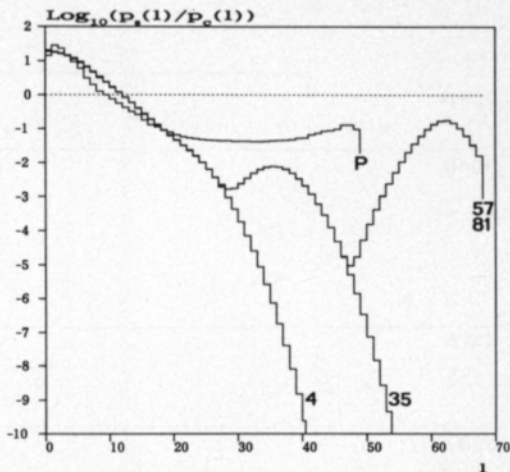


Fig. 18. Ratio of seed number probabilities in population 039A.

The preliminary function based on pilot crosses is marked by P, and the convergence results of different iterations are labelled with their iteration cycle numbers: 4, 35, 57 or 81. In the final study, function 57 was chosen to be used in the partitioning.

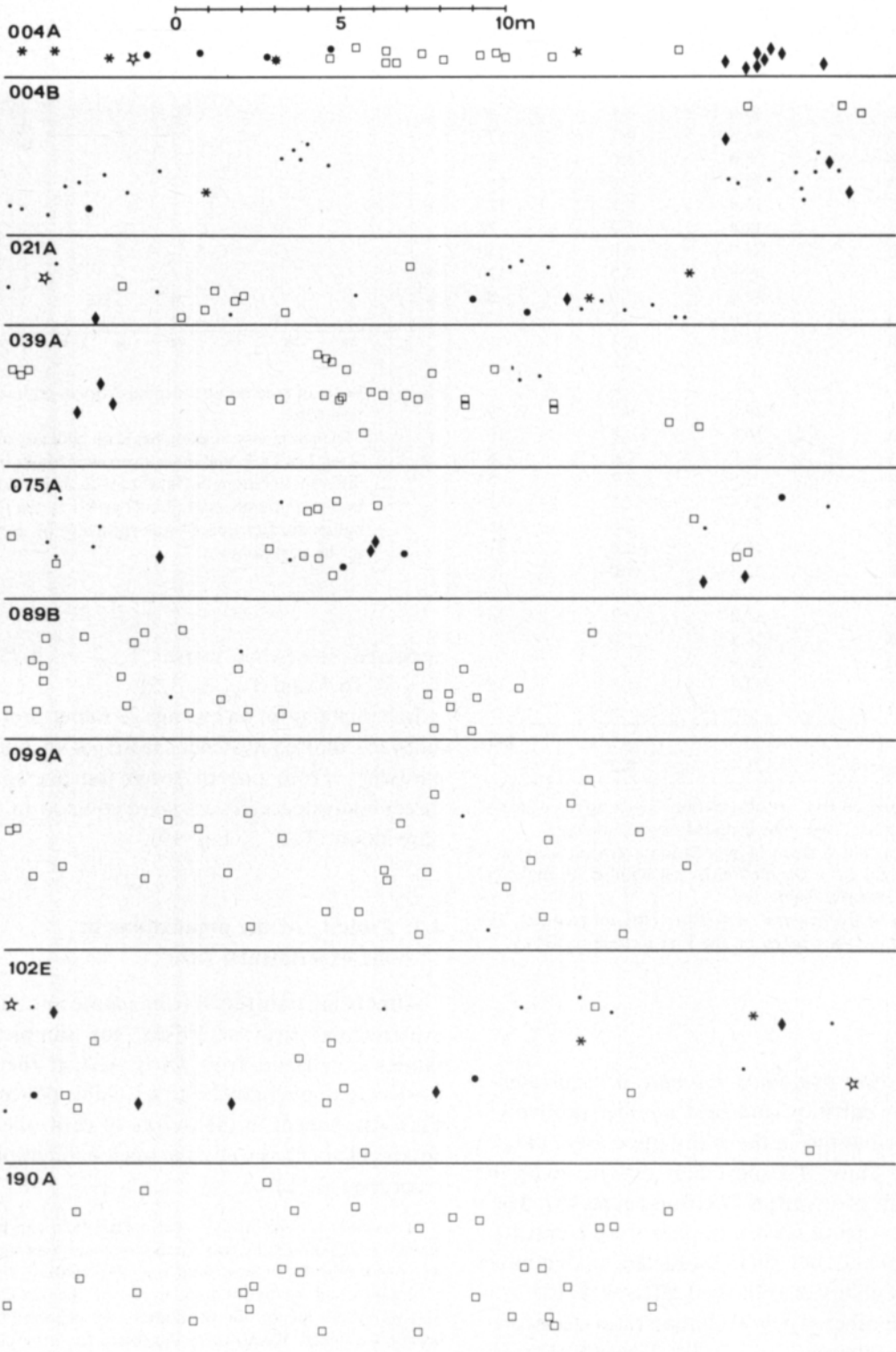
appeared somewhat worse:  $T_{81,35} = 4.75$ ,  $T_{81,4} = 16.7$  and  $T_{35,4} = 3.51$ .

In a population, on average 23 ramets were analyzed into equivalence classes of incompatibility. From one to seven (on average three) equivalence classes were counted in a population (Tab. 5, Fig. 19).

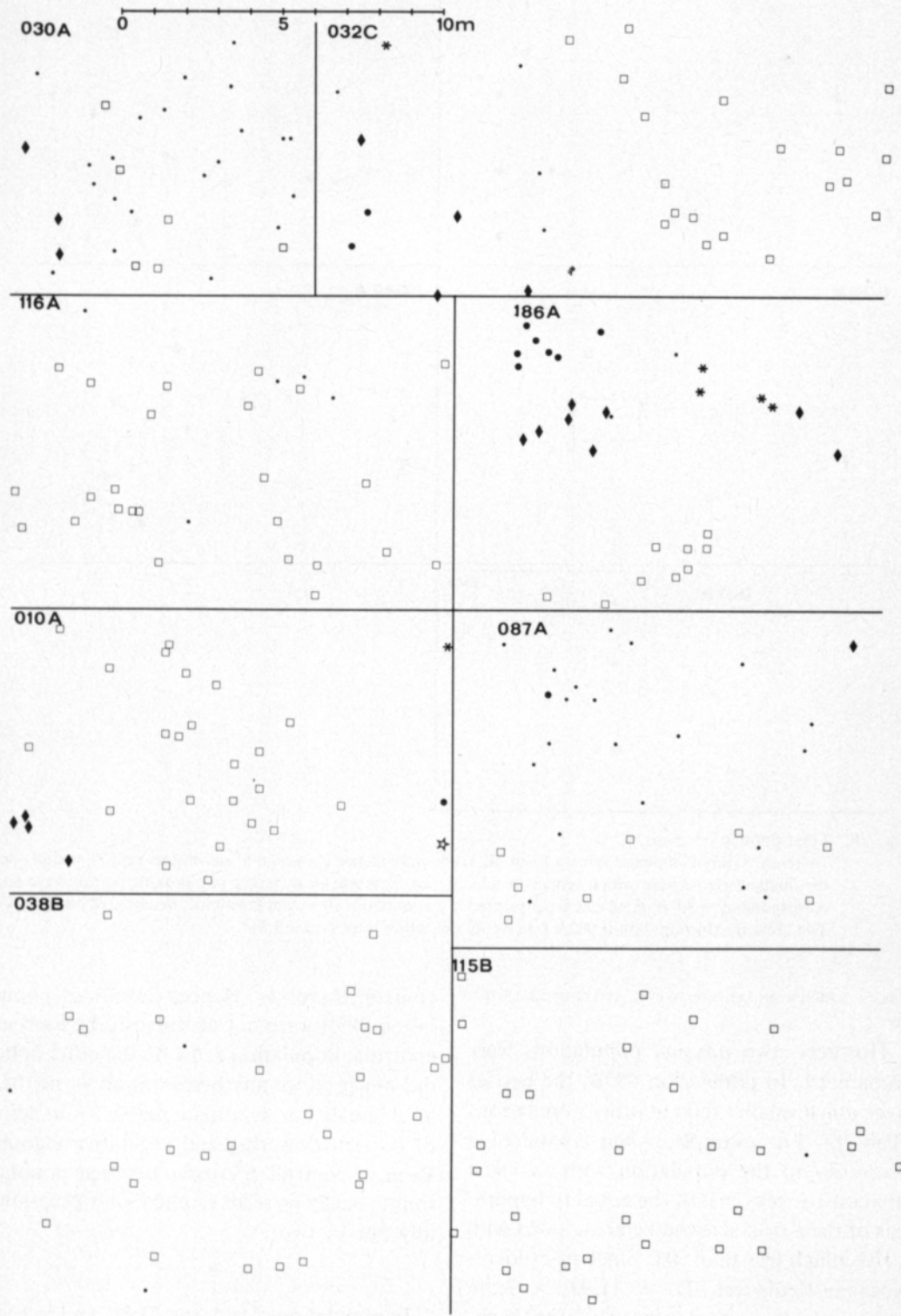
#### 4.2. Fruiting of the populations in the experimental field

After being transferred from nature into the experimental field at Viikki, the sampled ramets usually set fruit fairly well, if their flowers were pollinated with a suitable pollen. This was evident in the results of controlled crosses (Tab. 7) as well as in open pollination records (Tab. 8).

This ability to fruit in the experimental field can be made to suit both of the two competing main hypotheses. According to the 'environmental' explanation, the general external conditions are identical for the populations at Viikki. Hence, the populations are expected to behave in a similar manner there. According to the 'population structural' explanation, in the experimental field there will be a rich mixture of different incompatibility







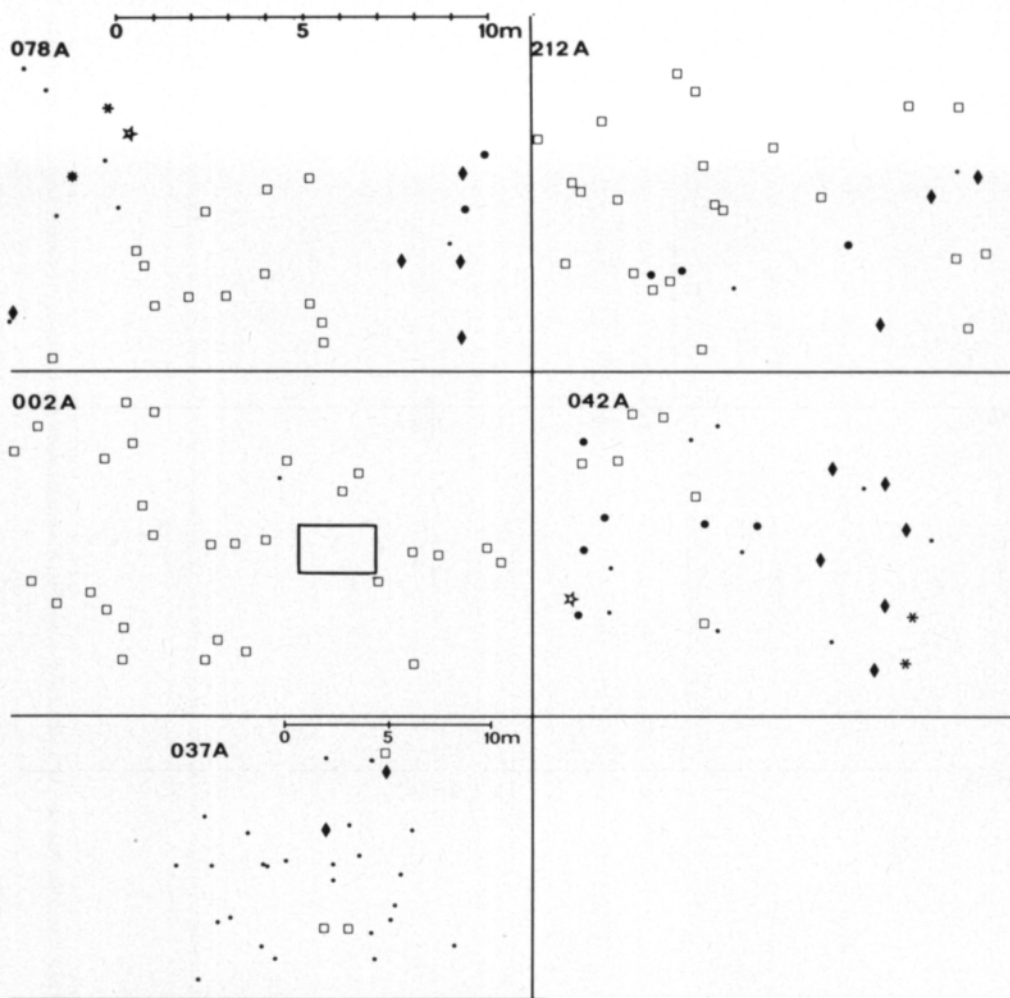


Fig. 19. Final population maps. Different symbols indicate ramets from different equivalence classes, and small dots represent those individuals in the sample which remained unanalyzed. Symbols in different populations do not have any correspondence. All populations are presented in a common scale except the almost unanalyzed one (037A). The rectangle in population 002A represents the 'alien' plot (see 4.3.8).

classes, which would ensure an even and abundant fruiting.

However, two *unusual* populations were recognized. In population 089B, the berries were much smaller than in other populations (Tab. 8). For example, when compared *a posteriori* to the population with the next smallest berries (204D), the equality hypothesis of their fruit size can be abandoned with a risk much less than .05. Since in Student-Newman-Keuls-test,  $D = 11.92 > 3.99$ , which is the 5% critical value. This small number of seeds in its berries was evident also in

controlled crosses. Hence, flowers of population 089B were not accepted to be used as controls. Population 116A on the other hand did not produce any berries at all — neither as a female nor as a male parent — in spite of its rich flowering and vegetative vigour. Even in controlled crosses between populations, usually no seeds resulted, only occasionally one or two.

In populations 116A and 089B, and in certain correspondents' accessions which only set

fruit poorly in the experimental field, cytological observations were made. Meiosis either in micro- or in megasporogenesis, as well as mitosis were studied. In divisions or chromosome number, nothing exceptional was recorded, except in one population. Namely, all ramets from population 116A proved to be *triploid* (Fig. 20, Fig. 3), which explains its totally fruitless condition.

No chromosomal misorders were recognized in population 089B. Its poor fruit set in any crosses could perhaps still be best explained by its general genetic background. Since, its flowers were smaller in diameter and contained fewer pistils than those of other populations. It also had fewer and more roundish petals than any other population under study (LARNA 1982). These features could be due eg. to an inbreeding depression — in a uniclass population, however, no reciprocal crosses could be made to test this. In principle, however, such features might also be a consequence of a “transferable” external factor common to all ramets in the population, say a (unnoticed) viral or fungal disease.

### 4.3. Structure and fruit set in natural populations

In statistical considerations, results from at most 22 populations were included. All populations in which less than five ramets had been analyzed into equivalence classes, were omitted (051A, 063A, 066A, 066D, 090A, 204D and 304A). The quantities at a population level, as well as the population averages of the

quantities at equivalence class or sample plot level (see below), are presented in Tab. 9.

#### 4.3.1. Observed quantities

a) At the population or incompatibility class level

FCI = *fruit number class* of the population; based on the average number of big berries (NBB) in the 30 sample plots divided by the average of cover percentage (C%); values: 0 (less than 1 fruit), 1 (1–15 fruits) or 2 (more than 15 fruits).

NEq = *number of equivalence classes* contained in (the sample taken from) the population; values: 1–7.

FZ = *fruiting zone* of the population (SAASTAMOINEN 1930); values: 1–4 (1 = optimal zone, see Fig. 5).

PAF = *proportion of alien flowers*; regarding a randomly taken flower, the average proportion of flowers capable of fertilizing it, ie. belonging to another incompatibility class; based on the accumulated number of past flowers (NPF); all flowers in a plot have been interpreted as belonging to the incompatibility class of the analyzed ramet.

EEq = *the areal extent of an equivalence class*; the average distance of an individual from the flowers of its own incompatibility class [m]; based on NPF.

b) At the plot or population level

NBB = *number of big berries*, ie. fruits containing at least 10 drupelets.

NSB = *number of small berries*, ie. fruits containing from 1 to 9 drupelets.

NUF = *number of unfertilized* (withering) past flowers.

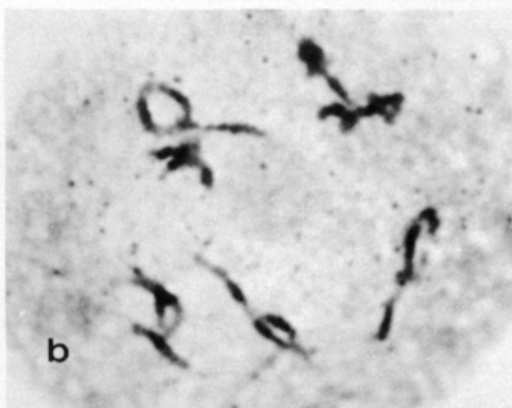
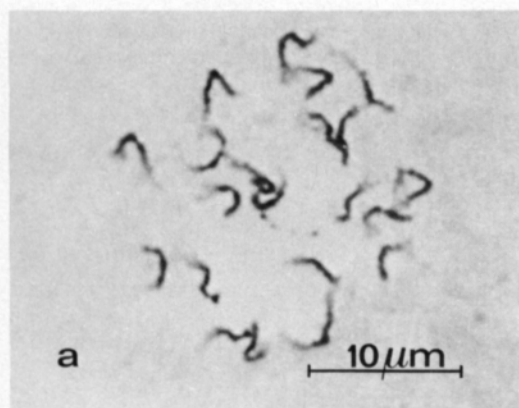


Fig. 20. Triploid ( $2n=21$ ) chromosomal constitution in arctic bramble population 116A.

a) Endomitotic prophase in a tapetum cell of individual n:o 16, in total 21 chromosomes can be counted, and b) diakinesis stage in the microsporogenesis of individual n:o 26, with 7 trivalents visible. (Photo: Maija Kotimäki).

- PBB = proportion of *big berries* ( $NBB/(NBB + NSB + NUF)$ )
- PSB = proportion of *small berries* ( $NSB/(NBB + NSB + NUF)$ )
- PUF = proportion of *unfertilized flowers* ( $NUF/(NBB + NSB + NUF)$ )
- NBu = number of *flower buds*
- NOF = number of *open flowers*
- NPF = accumulated number of *past flowers* ( $NBB + NSB + NUF + \text{unknown}$ )
- NF = accumulated number of *flowers* ( $NPF + NOF + NBu$ ); "fertility"
- C% = *cover percentage* of arctic bramble [%]
- DEq = average *distance* of an individual to *equivalent flowers* (ie. flowers of its own incompatibility class) [m]; based on NPF.
- DNEq = average *distance* of an individual to *non-equivalent flowers* (ie. flowers of other incompatibility classes) [m]; based on NPF.
- BBS =  $NBB/C\% =$  number of *big berries per square metre* of arctic bramble vegetation, corrected to a 100% cover.

c) At the plot or incompatibility class level

- DNA = distance to *nearest alien pollen parent*, ie. belonging to another incompatibility class [m]
- IN = "*index of neighbouring*", which counts the occasions when the plot in question serves as the nearest alien pollen parent for another plot.

#### 4.3.2. Population averages according to number of incompatibility classes

In the sample (of size 30 plots) taken, there occurred from one to seven equivalence classes in a population. The populations were *grouped* accordingly (Tab. 10). With regard to quite many quantities, the *means* of these groups *deviate* from each other statistically; ie. the hypothesis  $H_0$  stating that all these group means were equal, can be abandoned with a fairly minute risk probability ( $P^{KW}$ ). This holds especially for PAF, FCI, NBB, NSB and EEq, and also to a certain degree for PBB, PSB and PUF — all of which are important quantities with regard to our study. In general, the variables listed above will also show a *strong* (positive or negative) *association* (correlation) with the number of equivalence classes (Tab. 10).

In particular, the group of *uniclassic* populations is *different* from the other groups.

First, its mean usually represents the extreme value among them. Secondly, leaving the uniclassic populations

out of the test for group means, will cause a great increase in the risk probabilities ( $P^{KW}$ ) (NSB constituting the sole exception). Furthermore, the association with the number of equivalence classes will usually grow essentially weaker. In addition to that, the risk probability when abandoning the hypothesis of no association, displays a great increase. Though some increase is to be expected since the number of populations will be lower.

A direct proof is achieved by testing the mean of the uniclassic populations against the mean of other groups *combined* ( $P^{FP}$ ). Tested so, when abandoning the equality of these two means, the risk probabilities proved to be even smaller than in the original, unpooled data (NSB once more constituting an exception). Now the differences between the groups are found to be convincing also with regard to relative frequencies of berries (PBB, PSB or PUF).

The average value of fruit number class (FCI) is much *smaller* in the group of uniclassic populations than in other populations ( $P^{FP}$ ). Among themselves, the other populations do *not differ* except at a very great risk probability ( $P^{KW}$ ). Hence, the *strong positive association* between fruit number class and number of equivalence classes in the populations, is essentially caused by the *uniclassic* populations. Accordingly, the equality of fruit number class averages in uniclassic versus other populations, can be abandoned at an *extremely low risk* (applying U-test).

The quantity BBS, ie. the number of big berries per square metre, proves to behave in a similar manner to FCI. Since these two quantities are, however, technically very closely related to each other, this similarity in behaviour does not essentially give us any more information.

In uniclassic populations, the proportion of big (PBB) as well as small (PSB) berries among past flowers, is clearly (and at a small risk) lower than in the other populations. Accordingly, their proportion of unfertilized past flowers (PUF) is essentially (and at a small risk) higher than in other populations. Hence, in uniclassic populations, relative as well as absolute *intensity of fruiting* is *lower* than in other populations.

Actually, these results do not tell us very much that is new, but however they strengthen

Table 9. Population averages.

Popula- tion	FCI	BBS	NBB	NSB	NUF	PBB	PSB	PUF	NOF	NPF	NF	C%	NEq	EEq	DNEq	PAF	FZ
002A	0	0.1	0.0	1.0	21.2	.001	.045	.955	2.7	23.2	28.4	30.8	1	4.95	—	.000	1
004A	2	61.9	3.4	6.2	—	.401	.599	—	21.7	9.7	45.4	6.4	7	1.84	10.23	.740	1
004B	1	0.4 <sup>a</sup>	0.1 <sup>a</sup>	0.3 <sup>a</sup>	—	.122 <sup>a</sup>	.878 <sup>a</sup>	—	10.6	0.4	17.8	11.6	4	1.53	13.59	.899	1
010A	2	1.2 <sup>a</sup>	0.1 <sup>a</sup>	0.6 <sup>a</sup>	—	.201 <sup>a</sup>	.816 <sup>a</sup>	—	13.2	0.7	17.0	29.9	5	3.15	7.25	.183	3
021A	2	42.4	8.3	3.8	23.4	.203	.106	.691	6.0	35.4	58.8	25.4	5	1.95	7.95	.863	1
030A	1	3.0	0.2	1.5	11.2	.046	.131	.823	0.3	14.3	15.5	14.7	2	2.45	4.12	.488	1
032C	2	30.3	4.0	3.9	4.3	.369	.314	.317	2.6	12.7	27.7	23.1	4	3.75	9.76	.583	2
037A	2	27.3	3.9	1.7	3.8	.439	.192	.369	0.0	9.4	9.5	16.6	2	3.29	6.17	.713	1
038B	0	0.1	0.0	2.5	—	.002	.998	—	8.5	2.5	21.2	24.0	1	6.43	—	.000	4
039A	1	12.0	0.8	2.4	—	.340	.660	—	14.8	3.2	26.5	5.2	2	4.66	8.80	.176	3
042A	2	40.4	2.6	2.4	2.3	.361	.309	.329	0.1	8.8	9.0	8.0	5	2.28	5.59	.651	2
075A	2	75.1	2.8	11.2	9.9	.168	.423	.409	1.2	23.9	25.2	5.2	3	6.33	7.02	.485	4
078A	1	7.3	0.7	3.4	8.6	.038	.316	.645	0.4	13.6	14.6	8.9	6	2.92	6.52	.503	1
087A	2	104.9	11.1	16.2	8.6	.377	.357	.266	0.9	35.9	37.3	10.6	3	3.96	7.13	.417	2
089B	0	0.0	0.0	3.5	143.7	.000	.031	.969	38.1	147.2	241.0	37.5	1	6.55	—	.000	2
099A	0	0.0 <sup>a</sup>	0.0 <sup>a</sup>	0.2 <sup>a</sup>	—	.000 <sup>a</sup>	1.000 <sup>a</sup>	—	23.3	0.2	38.1	32.1	1	6.98	—	.000	3
102E	0	0.0	0.0	0.2	8.4	.000	.046	.954	0.9	8.7	18.6	11.3	5	8.95	9.75	.634	1
115B	0	1.7	0.0	0.3	9.3	.019	.032	.948	2.2	10.4	14.8	7.7	1	6.13	—	.000	1
116A	0	0.0	0.0	0.0	17.9	.000	.001	.999	2.9	17.9	22.3	7.1	1	5.89	—	.000	3
186A	2	40.2	3.3	5.7	10.5	.174	.282	.545	0.0	20.8	21.0	9.3	4	1.90	5.49	.756	3
190A	0	1.1	0.0	1.6	5.5	.037	.180	.783	4.2	7.2	21.5	7.5	1	6.28	—	.000	4
212A	2	28.6	17.7	9.0	26.1	.340	.147	.513	22.2	52.9	125.0	58.7	3	4.44	5.94	.319	3

<sup>a</sup> The records were made so early in the summer as to render them uninformative.

Table 10. Population averages according to number of equivalence classes

	Number of equivalence classes (NEq)										Correlation <sup>a</sup> with NEq	
	1	2	3	4	5	6—7	P <sup>KW</sup>	P <sup>ekw</sup>	2—7	P <sup>FP</sup>	$\tau$	P
FCI	0.00	1.33	2.00	1.67	1.50	1.50	<b>.0011</b>	.70	1.60	<b>.00005<sup>U</sup></b>	.44	<b>.0016</b>
											.03 <sup>c</sup>	.42
BBS	0.50	14.1	69.5	35.3	27.6	34.6	<i>.013</i>	.40	36.4	<b>.0002</b>	.48	<b>.0012</b>
											.18 <sup>c</sup>	.18
NBB	0.00	1.63	10.5	3.65	3.63	2.05	<b>.0035</b>	.47	4.52	<b>.0003</b>	.40	<b>.0064</b>
											— .05 <sup>c</sup>	.57
NSB	1.48	1.87	12.1	4.80	2.13	4.80	<b>.0046</b>	<b>.0047</b>	5.20	<i>.012</i>	.29	<i>.040</i>
											.01 <sup>c</sup>	.43
NUF	39.5	7.50	14.9	7.40		10.7	.68	.78	10.6	<i>.079</i>	— .23	.13
											— .11	.38
PBB	.011	.243	.295	.272		.151	<i>.036</i>	.68	.229	<b>.0021</b>	.27	<i>.070</i>
											— .27 <sup>c</sup>	.86
PSB	.058	.162	.309	.298		.194	<i>.035</i>	.45	.238	<b>.0035</b>	.38	<i>.016</i>
											— .02 <sup>c</sup>	.50
PUF	.931	.596	.396	.431		.655	<i>.019</i>	.49	.533	<b>.0016</b>	— .32	<i>.048</i>
											.20 <sup>2</sup>	.82
NOF	11.7	5.03	8.10	4.40	5.05	11.1	.71	.88	6.33	.86	— .14	.81
											.11 <sup>c</sup>	.25
NPF	41.2	11.9	37.6	16.8		16.6	.31	.14	21.5	.75	— .08	.66
											— .20 <sup>c</sup>	.78
NF	65.6	12.5	62.5	24.4		25.3	.19	.19	32.9	.80	— .15	.77
											— .09 <sup>c</sup>	.62
C%	21.0	12.2	24.8	14.7	18.7	7.65	.79	.71	16.3	.76	— .12	.77
											— .07 <sup>c</sup>	.61
EEq	—	3.47	4.91	2.39	4.08	2.38	—	.20	3.56	—	— .27 <sup>c</sup>	.10
									7.73 <sup>3</sup>	.61		
DNEq	—	6.36	6.70	9.61	7.64	8.38	—	.73	8.02 <sup>3</sup>	.14	.23 <sup>c</sup>	.90
PAF	—	.459	.407	.670	.716	.622	—	.14	.564	—	.38 <sup>c</sup>	.029
FZ	2.57	1.67	3.00	2.00	1.75	1.00	.32	.28	1.93	.15 <sup>U</sup>	— .23	.072
											— .21 <sup>c</sup>	.16

Risk probabilities below .05 or (in a one-tailed test) over .95 are presented in *italics*, and those below .01 or (in a one-tailed test) over .99 in **bold face** type.

<sup>FP</sup> One-tailed Fisher-Pitman test. The group of uniclassic populations is compared with the combined group of populations with 2—7 classes. The direction of the counter-hypothesis  $H_1$  is the same as in Table 13.

<sup>U</sup> Mann-Whitney U-test.

<sup>KW</sup> Kruskal-Wallis variance analysis for ranks.

<sup>3</sup> Populations with 3—7 classes pooled (and compared with the group of populations with 2 classes).

<sup>a</sup> See Tab. 13.

<sup>c</sup> Uniclassic populations excluded.

one of the basic assumptions for the study: in arctic bramble, self-incompatibility is quite *strict*. They also provide indirect evidence that the analytic *tools* constructed for the study were functioning properly as well as that the *sample size* (30 ramets) was adequate. Since, all populations which were estimated as

uniclassic (by analyzing the sample with the crossing system), proved to have been fruitless ones in nature.

Noticeably, the average fruiting zone (FZ) in populations did not differ between these groups which were based on the number of equivalence classes. Hence, populations with different numbers of equivalence classes were distributed in a *similar* way into the different fruiting zones.

It should also be noted that the number of equivalence classes seems to have no apparent numerical relationship to such 'vigour' quantities as accumulated number of flowers (NF) or cover percentage of arctic bramble (C%). Somewhat unexpectedly, this also holds true with regard to the average distance of an individual to non-equivalent flowers (DNEq).

Considering the group of *uniclassic* populations, the proportion of alien flowers (PAF) would be a *trivial* quantity. Namely, it tells nothing more than is implied already by the feature of being uniclassic. Since, in uniclassic populations, any 'alien' flowers will certainly not be available. Furthermore, with regard to the uniclassic populations, the areal extent of an incompatibility class (EEq) would contain a *systematic error*. Since, populations smaller in dimension than a specified minimum, were excluded from the study. Thus, no equivalence classes with a *small area* are represented in the uniclassic populations. Concerning the present grouping, therefore, the group of uniclassic populations was excluded from the statistical tests with these two variables.

#### 4.3.3. Population averages according to fruit number class

The populations were arranged into three groups according to fruit number class (FCI) (Tab. 11). For roughly the same variables as above — variables which were *central* to the problem being studied — the group averages *differed* from each other at a low level of risk.

The number of equivalence classes in a population (NEq) was recorded to definitely *increase* with fruit number class (this strong positive correlation will, however, disappear totally, if uniclassic populations are removed from the data). In *non-fruiting* populations, there are on average much *fewer equivalence classes* than in other populations, and the hypothesis of equality can be abandoned at a low risk (P<sup>FP</sup>). Also the average proportion of alien flowers in the population (PAF) (which now is a non-trivial quantity in all groups), was *very much smaller* in the group of non-fruiting populations than in other populations. Additionally, at a fairly low risk level, the proportion of alien flowers was higher in richly fruiting populations than in the populations setting fruit poorly. There was a *strong positive association* of the proportion of alien flowers with fruit number class (this however disappears if uniclassic populations are excluded).

The grouping variable (FCI) will be closely related to fruit set quantities, especially the number of big berries (BBS or NBB). Therefore, the strong differences between group averages in *fruit set* quantities, will be mostly *technical* in nature. Hence, they will yield essentially no extra information for us. The same applies for the ostensibly strong associations of fruit set quantities with fruit number class.

In relevant data (where uniclassic populations have been removed), the areal extent of an equivalence class (EEq) showed neither differences between fruit number classes nor a correlation with fruit number class but at a high risk probability. Quantities measuring flower numbers, cover percentage or distance to other equivalence classes (NOF, NPF, C%, DNEq) showed no correlation with fruit number class. According to fruit number classes, fruiting zone averages (FZ) do not differ but at a high risk, and no signs of an association between these two quantities can be seen.

The results of NEq and PAF constitutes a considerable evidence *against* the non-genetic or 'environmental' explanations for absence of fruit set, but in favour of the explanation based on incompatibility class structure in the populations. If the inability of a population to set fruit in nature were decisively caused by an *external* factor (eg. temperature, moisture, soil constitution, frost, lack of pollinating insects, or a weak vigour of the clones), one could hardly expect to get such a strong correlation of the number of equivalence classes (NEq) or the proportion of alien flowers (PAF) with fruit number class (FCI). On the contrary, on the basis of the central hypothesis of the present study — stating that availability of genetically successful (compatible) pollen is the most essential factor — just such interdependences could be expected.

#### 4.3.4. Population averages according to fruiting zone

The populations were classified in four groups according to their fruiting zone (Tab. 12). Unlike the previous groupings, the group averages do *not differ* from each other but at a high risk probability. This will apply to any quantity under study.

Hence, considering eg. their fruit number class averages (FCI), the groups based on fruiting zone (FZ) do not differ from each other but at a great risk. Nor is there any evidence

Table 11. Population averages according to fruit number class

	Fruit number class (FCI)						Correlation <sup>a</sup> with FCI	
	0	1	2	<i>P</i> <sup>KW</sup>	1—2	<i>P</i> <sup>FP</sup>	$\tau$	<i>P</i>
BBS	0.43	7.43	50.1	< .00001	39.5	.00001	.65 .50 <sup>c</sup>	< .0001 .0051
NBB	0.00	0.57	6.34	< .00001	4.90	.00002	.65 .50 <sup>c</sup>	< .0001 .0051
NSB	1.30	2.43	6.68	.0021	5.62	.0016	.49 .41 <sup>c</sup>	.0009 .021
NUF	34.3	9.90	11.1	.73	10.9	.13	-.14 .02 <sup>c</sup>	.25 .56
PBB	.009	.042	.304	< .00001	.252	.0001	.63 .47 <sup>c</sup>	.0001 .013
PSB	.056	.224	.266	.0041	.258	.0005	.47 .22 <sup>c</sup>	.0043 .14
PUF	.935	.734	.430	.00007	.491	.0002	-.60 -.44 <sup>c</sup>	.0004 .043
NOF	10.4	6.53	6.79	.50	6.71	.78	-.16 .01 <sup>c</sup>	.84 .46
NPF	35.8	14.0	25.0	.68	22.8	.66	.15 .25 <sup>c</sup>	.20 .11
NF	57.8	15.1	39.2	.40	34.4	.72	.00 .18 <sup>c</sup>	.48 .18
C%	19.8	10.1	19.3	.60	16.7	.69	-.03 .12 <sup>c</sup>	.57 .25
NEq	1.50	3.50	4.10	.0021	3.93	.0015	.44 .03 <sup>c</sup>	.0016 .42
EEq	6.52	2.89	3.29	.0002	3.18 4.10 <sup>1</sup>	.00002 .24	-.08 <sup>c</sup>	.39
DNEq		8.56	7.25		8.56 <sup>1</sup>	.18	-.11 <sup>c</sup>	.31
PAF	.091	.389	.614	.0019	.558	.0004 .051 <sup>12</sup>	.50 .17 <sup>c</sup>	.0009 .18
FZ	2.38	1.50	2.20	.44	2.00	.29 <sup>U</sup>	-.01 .23 <sup>c</sup>	.48 .90

Risk probabilities below .05 or (in a one-tailed test) over .95 are presented in *italics*, and those below .01 or (in a one-tailed test) over .99 in **bold face** type.

<sup>1</sup> Fruit number classes 0 and 1 pooled.

<sup>12</sup> Classes 1 and 2 compared.

<sup>FP</sup> One-tailed Fisher-Pitman test. The group of non-fruiting populations is compared with the combined group of populations with poor or rich fruit set. The direction of the counter-hypothesis  $H_1$  is the same as in Table 13.

<sup>U</sup> Mann-Whitney U-test.

<sup>KW</sup> Kruskal-Wallis variance analysis for ranks.

<sup>a</sup> See Tab. 13.

<sup>c</sup> Uniclassic populations excluded.

to suggest an association between these two quantities. Thus, the *sampling* of the studied populations from all over Finland, seems to have *succeeded* in the desired manner.

The original aim was at a *stratified* sampling with regard to the fruiting zones. Richly fruiting, poorly fruiting and non-fruiting populations should have been acquired in proportions identical in each zone. In practice, difficulties arose. Since, due to imprecise or even misleading in-



Table 12. Population averages according to fruiting zone.

	Fruiting zone (FZ)								Correlation <sup>a</sup> with FZ	
	1	2	3	4	P <sup>KW</sup>	1—2	3—4	P <sup>FP</sup>	$\tau$	P
FCI	1.00	1.50	1.17	0.67	.69	1.15	1.00	.44 <sup>U</sup>	— .01	.48
									.23 <sup>e</sup>	.90
BBS	18.0	43.9	20.2	25.4	.87	26.6	22.4	.41	.01	.55
									.22 <sup>e</sup>	.87
NBB	2.06	4.43	5.45	0.93	.56	2.85	3.51	.62	— .02	.47
									.17 <sup>e</sup>	.82
NSB	2.26	6.50	4.28	5.10	.41	3.68	4.63	.68	.19	.88
									.38 <sup>e</sup>	.971
NUF	12.3	39.7	16.1		.39	22.3	14.0	.43	.02	.45
									.11 <sup>e</sup>	.27
PBB	.107	.277	.171	.103	.72	.168	.144	.39	.00	.52
									.05 <sup>e</sup>	.62
PSB	.124	.253	.143	.302	.39	.171	.207	.69	.16	.82
									.35 <sup>e</sup>	.94
PUF	.769	.470	.686	.596	.43	.661	.650	.53	— .11	.69
									— .31 <sup>e</sup>	.89
NOF	4.98	10.4	12.7	4.63	.48	6.65	10.0	.77	.16	.86
									.12 <sup>e</sup>	.75
NPF	16.4	51.2	30.5	15.6	.56	29.1	24.5	.57	.13	.77
									.31 <sup>e</sup>	.92
NF	22.9	78.8	56.1	23.4	.59	43.2	43.0	.58	.18	.85
									.24 <sup>e</sup>	.86
C%	14.8	19.8	23.7	12.2	.70	16.4	19.9	.72	— .06	.37
									— .10 <sup>e</sup>	.35
NEq	3.67	3.25	2.67	1.67	.44	3.54	2.33	.089	— .23	.072
									— .21 <sup>e</sup>	.16
EEq	3.78	4.14	4.50	6.35	.38 <sup>e</sup>	3.29 <sup>e</sup>	4.10 <sup>e</sup>	.24	.31 <sup>e</sup>	.046
DNEq	8.33	7.49	6.90		.68 <sup>e</sup>	8.08 <sup>e</sup>	6.90 <sup>e</sup>	.80	— .16 <sup>e</sup>	.78
PAF	.493	.413	.313	.162	.42	.466	.248	.073	— .26	.062
						.621 <sup>e</sup>	.434 <sup>e</sup>	.062	— .32 <sup>e</sup>	.082

Risk probabilities below .05 or (in a one-tailed test) over .95 are presented in *italics*, and those below .01 or (in a one-tailed test) over .99 in **bold face** type.

<sup>FP</sup> One-tailed Fisher-Pitman test. The direction of the counter-hypothesis  $H_1$  is the same as in Table 13.

<sup>U</sup> Mann-Whitney U-test.

<sup>KW</sup> Kruskal-Wallis variance analysis for ranks.

<sup>a</sup> See Tab. 13.

<sup>e</sup> Uniclassic populations excluded.

formation received from the arctic bramble survey, many of the populations originally chosen for the study had to be changed to other ones or reclassified with regard to their actual fruit set (see 2.1. and 2.3). It should be noted that due to such stratified sampling, *no information* whatsoever could be attained concerning the question of putative *differences in frequency* of the different fruit number classes in different zones.

*Nor* does the average number of equivalence classes (NEq) in populations differ in different zones except at a great risk probability. However some signs of a slight

negative association between NEq and fruiting zone, and also between the proportion of alien flowers (PAF) and FZ, may be seen. That is — if the sampling is made in the way it was done in the present study — the populations sampled from the more 'optimal' zones will on average contain no more (or only slightly more) equivalence classes or alien flowers than the populations from other zones.

In the data fraction, where uniclassic populations are outruled, the number of small berries (NSB) and areal extent of an equivalence class on average in the popula-

tion (EEq), show a positive association with fruiting zone number at a fairly small risk. Hence, regarding a common fruiting class and populations containing more than one incompatibility class, then on average smaller-sized equivalence classes and fewer small berries will be encountered in the more productive zones. This has however no implications with regard to the competing two main hypotheses.

#### 4.3.5. Correlations of population averages

The pairwise associations (correlations) of all quantities at a population level have been combined into a common table (13). The correlations of the three *central* quantities (NEq, FCI and FZ) with other ones have already been considered (see above and Tables 10, 11 and 12).

The risk probabilities presented (Tab. 13), refer to the null hypothesis  $H_0$ : there is no correlation. A one-sided test is chosen, and depending on the pair of variables considered, the counter-hypothesis will be *either*  $H_1$ : a positive correlation exists *or*  $H_1$ : a negative correlation is postulated.

In each case, the counter-hypothesis  $H_1$  has been fixed beforehand on the basis of theoretical postulates considering these kinds of populations or their structure. Considering the central variables in this study, the counter hypothesis  $H_1$  has usually been determined directly on the basis of the *basic hypothesis* in the study. This states that with regard to the fruiting of vigorous arctic bramble populations, the most essential factor is the availability of compatible pollen (see 1.3). If this basic hypothesis is true, a *positive* correlation should prevail between eg. fruit set and number of equivalence classes.

The average proportion of alien flowers in a population (PAF) shows a *negative* correlation with the proportion of unfertilized past flowers (PUF) at a fairly low risk, and a *positive* correlation with the proportion of big berries (PBB). The more equivalence classes (NEq) there are in a population, the greater is the average proportion of alien flowers (PAF) — at a very low risk. At a fairly low risk, this applies also for a data, where uniclassic populations have been excluded. In

addition, the quantity PAF expresses a *strong positive* association with fruit number class (FCI), number of big berries (NBB) and number of big berries per square metre (BBS).

In addition, an array of other such strong correlations of an important population structural variable with central fruiting variables, have been presented above (4.3.2, 4.3.3, Tables 10 and 11). Without any *a posteriori* “rescue” hypotheses, these correlations could *not* have been expected on the basis of the *environmental* explanation for fruitless populations. On the contrary, on the basis of the *population structural* explanation, such correlations are to be expected. Furthermore, all the correlations occurred in the direction (ie. positive or negative) predicted by the latter hypothesis. Therefore, correlation results give *support* to our *main hypothesis*, namely that the lack of compatible flowers is the most important factor, in causing missing fruit set in vigorous arctic bramble populations.

For a general overview, a complete table of associations has been presented. Though, many of the correlations will not carry much information concerning the two competing main hypotheses. With regard to such *less important* pairs of variables, in place of or in addition to the basic hypothesis, other ecological assumptions have proved necessary in fixing the counter hypothesis  $H_1$ . Often one of the following *extra assumptions* has been used.

a) The incompatibility classes in the population (do constitute different pure clones and) are clearly *separate* from each other spatially. Thence, as a result of an increase eg. in the areal extent of the equivalence classes, or in the number or density of their flowers, the proportion of within-class pollinations would *increase* (see TAMMISOLA 1986).

b) In considerations concerning fruiting zones (FZ), the zonation introduced by SAASTAMOINEN (1930) is postulated to hold in nature.

c) A plot is supposed to produce more flowers the more distant it is from other plots, especially from plots of its own incompatibility class. In the first instance, this is based on the radial expansion of an arctic bramble genet — in the centre, old and degenerating ramets are met, while at the margins (ie. on average most distant from their own class) younger and more vigorous ramets will occur.

The proportion of unfertilized, withering past flowers (PUF) shows a *strong negative* association with any other fruiting quantity (FCI, BBS, NBB, NSB, PBB or PSB). These associations will however *not* provide us with any essential extra information for a choice between the two rivaling basic hypotheses. Since, according to each of them, withering flowers will not develop into berries. In

Table 13. Correlations of population averages. Lower number is Kendall's rank correlation coefficient ( $\tau$ ), upper number is the risk probability in a one-tailed test.

	FCI	FCF	BBS	NBB	NSB	NUFB	PBB	PSB	PUF	NOF	NPF	NF	C%	NEq	NEq <sup>c</sup>	EEq <sup>c</sup>	DNEq	PAF	PAF <sup>e</sup>	FZ	FZ <sup>e</sup>	
FCI	•																					
FCF		•																				
BBS			•																			
NBB				•																		
NSB					•																	
NUF						•																
PBB							•															
PSB								•														
PUF									•													
NOF										•												
NPF											•											
NF												•										
C%													•									
NEq														•								
NEq <sup>c</sup>															•							
EEq <sup>c</sup>																•						
DNEq																	•					
PAF																		•				
PAF <sup>e</sup>																			•			
FZ																				•		
FZ <sup>e</sup>																					•	

H<sub>0</sub>: no correlation; in the upper triangle, H<sub>1</sub> dictates a positive correlation, in the lower a negative correlation. The choice of the counter-hypothesis H<sub>1</sub> is explained in the text.

Risk probabilities below .05 or (in a one-tailed test) over .95 are presented in *italics*, and those below .01 or (in a one-tailed test) over .99 in **bold face** type.

<sup>c</sup> Uniclassic populations excluded.

<sup>e</sup> In a two-tailed test.

addition, the proportions of big (PBB) or small (PSB) berries express a *strong positive* association with each other. Also this result is indifferent in regard to the two competing basic explanations.

Somewhat unexpectedly, the average distance of an individual from non-equivalent flowers (DNEq) did not show any associations with any other primary variable, but only with the average number of open flowers (NOF). Even this single association may be due to pure chance. Since, the number of open flowers will vary "erratically" from one population to another. Namely, the collection journeys to different populations were made at widely different times — and for example during late summer, only very few flowers are present.

Some of the obvious associations in Table 13 are to be regarded as *trivial*. Namely, if the quantities in question contain common components (as eg. NF and NOF or NPF), they will express a "technical" association.

It should be noted that the correlations presented are *non-parametric*. As an association measure, Kendall's rank correlation coefficient ( $\tau$ ) was chosen. Though it accepts values from  $-1$  to  $+1$ , its metrics (underlying scale) is *not identical* to that of an ordinary product-moment correlation coefficient by Pearson. The values of  $\tau$  are as a rule smaller in absolute value than the respective values obtained using product-moment correlation coefficient. Hence their numerical values indicate a *stronger* level of association than an equal value by the ordinary correlation coefficient.

#### 4.3.6. *Quantities within a population according to equivalence class*

Four populations with the most equivalence classes and least unclassified ramets, were chosen for consideration. Two of them had set fruit richly (004A and 186A), one poorly (078A) and one not at all (102E). In each population separately, the equality of the mean values of its equivalence classes was tested applying Kruskal-Wallis variance analysis for rank values (Tab. 14).

In population 004A, the equivalence classes differ from each other at a small risk with regard to almost any variable. In population 186A, the equivalence classes generally differ with regard to population structural variables — and at a particularly low risk with regard to the distance from nearest alien pollen parent (DNA) and average distance from other incompatibility classes (DNEq). Though, considering fruiting variables such differences do not occur other than at a great risk probability. The general view of population 078A is

similar to that in 004A, ie. equivalence class averages generally differ. Regarding the fruiting variables, however, the risk when abandoning the equality hypothesis is as a rule greater than in 004A — which can be expected on the basis of the poor fruit set. In population 102E, almost no differences at all exist between the equivalence class averages, except at a great risk. Though, its classes still differ at a low risk in regard to a single population structural quantity, ie. "index of neighbouring" (IN).

Remarkably, the quantity on which fruit number classification of populations is based, ie. BBS proves to be always fairly *homogeneous* between equivalence classes in a population. Also other *fruiting variables* seem to be less efficient in comparing equivalence classes than they were in comparing populations.

In part this may be solely a technical matter, since the equivalence class averages have a greater mean error than the population averages. Hence, though there were real differences between equivalence classes, the risk probability would still remain fairly high due to too few plots in the classes. There is of course another possible explanation, stating that in a population, only small or no differences at all occur between the average fruit set in different equivalence classes.

Regarding *population structural* quantities, the equivalence class averages in a population often *differ* from each other at a low risk.

However, in different populations quite *different* structural quantities proved inefficient in this respect (IN in 186A, DNEq in 078A, DNA and DNEq in 102E). Hence, there presumably is something different in the equivalence class structure in different populations.

A general view is attained that in a population, there usually exists some structural differences between the equivalence classes. These differences however will not in all cases imply any (convincing) differences in fruit set. Thus, in certain populations, the possibility for a flower to become fertilized may be essentially *equal* overall the population. Or at least so similar that in order to achieve conclusive evidence for differences in fruit set according to equivalence classes, a much larger sample should have been taken.

Table 14. Comparisons of equivalence class averages made within each of four populations.

	Population averages				P <sup>KW</sup> risk when rejecting identity of equivalence class averages			
	004A	186A	078A	102E	004A	186A	078A	102E
BBS	61.9	40.2	7.3	0.0	.17	.11	.18	1.00
NBB	3.4	3.3	.67	.00	.016	.073	.19	1.00
NSB	6.2	5.7	3.4	.23	.0004	.81	.058	.26
NUF	—	10.5	8.6	8.4	—	.040	.033	.36
PBB	.401	.174	.038	.000	.0032	.34	.017	1.00
PSB	.599	.281	.316	.048	.0033	.39	.43	.10
PUF	—	.545	.645	.952	—	.63	.69	.10
-----								
NOF	21.7	.03	0.4	0.9	.0001	1.00	.019	.56
NPF	9.7	20.8	13.6	8.7	.0033	.21	.023	.30
NF	45.4	21.0	14.6	18.6	.0032	.24	.022	.40
C%	6.4	9.3	8.9	11.3	.68	.27	.37	.9995
-----								
EEq	1.84	1.90	2.92	8.95	<.00001	.013	.0003	.094
DNEq	10.2	5.5	6.5	9.8	<.00001	.0005	.64	.21
DNA	1.85	2.86	2.77	1.84	.0053	.00002	.00003	.21
IN	1	1	1	1	.0060	.10	.0031	.0093

<sup>KW</sup> Kruskal-Wallis variance analysis for ranks.

#### 4.3.7. Correlations of quantities within a population

In the four populations (004A, 186A, 078A and 102E), pairwise associations of the variables were calculated at the plot level. The most relevant ones with regard to our competing two main hypotheses, will be considered.

In **population 004A** (Tab. 15), observations concerning unfertilized flowers (PUF) are missing. Therefore, the proportion of small berries (PSB) has adopted a mixed role — it represents in part also PUF. The proportion of big berries (PBB) is in a strong *negative* correlation with the average distance from non-equivalent flowers (DNEq) and (less strongly) with the distance to nearest alien pollen parent (DNA). A *positive* association is shown by PBB with the variable DEq, ie. the average distance of a ramet from the flowers of its own equivalence class.

The quantities DNEq and DNA express a strong *positive* association with each other. This is to be expected, since essentially they roughly measure the same character. On the other hand, DNA shows a strong *negative* association with the index of neighbouring (IN). This is just as expected, since a plot situated uncommonly far from its nearest alien plot cannot in general constitute the

nearest alien plot to many other plots. The average distance of an individual from equivalent (DEq) and non-equivalent (DNEq) flowers, are in a *negative* association with each other. On the basis of the extra hypothesis which postulates that classes were spatially separate, such an association is to be expected.

In **population 186A** (Tab. 16), the situation appears to be quite dissimilar. Associations *between* population structural and fruiting variables — which are interesting with regard to the competing two main hypotheses — can certainly *not* be found any more. DEq is the only structural quantity still yielding any signs of such an association. Even with this sole exception, its association with PUF is just in the *opposite* direction from the expected one. Since, the further the plot is situated from its own class, the greater proportion of its flowers will wither.

Unlike population 004A, DEq is now in a strong (negative) association with cover percentage (C%). Furthermore, no association exists between DNEq and DEq, which in 004A displayed a negative but slight association. The relationships between structural variables DNEq, DNA and IN, have however remained similar as in population 004A.

In **population 078A** (Tab. 17), there exists only a single strong correlation of a population structural quantity with any fruiting vari-

Table 15. Correlations of quantities in population 004A. Lower number is Kendall's rank correlation coefficient ( $\tau$ ), upper number is the risk probability in a one-tailed test.

	BBS	NBB	PBB	PSB	NOF	NF	C%	DEq	DNEq	DNA	IN
BBS	•	.0002 <b>.44</b>	.0073 <b>.31</b>		.91 -.17	.968 -.24		.19 .11			.063 .20
NBB		•	.0033 <b>.35</b>		.25 .08	.15 .13	.0065 <b>.31</b>	.23 .09			.21 .10
PBB			•					.014 .29			.25 .09
PSB	.0090 -.31	.0037 -.35	<.0001 -.96	•	.0041 .34	.044 .22	.83 -.13		.0003 <b>.44</b>	.022 .26	
NOF			.0046 -.34		•	<.0001 .78	.0014 .37	.9903 -.30	.012 .29	.040 .22	
NF			.052 -.22			•	<.0001 <b>.54</b>	.946 -.21	.13 .14	.38 .04	
C%	.040 -.23		.84 .13				•	.60 -.03	.88 -.16	.912 -.18	
DEq				.015 -.29				•			.32 .06
DNEq	.16 -.13	.14 -.15	.0004 -.44					.043 -.23	•	.0009 <b>.39</b>	
DNA	.23 -.10	.094 -.17	.026 -.26					.17 -.13		•	
IN				.26 -.09	.077 -.19	.21 -.11	.40 -.03		.12 -.16	.0006 -.41	•

$H_0$ : no correlation; in the *upper* triangle,  $H_1$  dictates a *positive* correlation, in the lower a *negative* correlation. The choice of the counter-hypothesis  $H_1$  is explained in the text.

Risk probabilities below .05 or over .95 are presented in *italics*, and those below .01 or over .99 in **bold face** type.

able. That is the association of DNEq (average distance from non-equivalent flowers) with PUF (proportion of unfertilized flowers). In population 004A (there with PSB), the respective association was strong and positive, while in population 186A it was almost non-existent. Now, in population 078A, it is *strong* but *negative*. Hence, the proportion of unfertilized flowers will *decrease*, the greater the distance from other equivalence classes!

On the other hand, in population 078A, *neither* PUF *nor* DNEq will differ in different equivalence classes except at a very high risk indeed (Tab. 14). In fact, when abandoning the equality of class averages, their risk is greater than that of any other variable. Hence, even an apparent association between these two variables has interpretative value only *within* an equivalence class.

In addition, there are some indications of a positive association of DNEq with PBB (proportion of big berries) and of a negative association of DEq (average distance from

equivalent flowers) with PBB.

The structure of population 078A seems to be *different* from that in the previous two other populations. Since, now the population structural variables DNEq and DNA (distance to nearest alien pollen parent) do *not* show any association with each other.

In **population 102E** (Tab. 18), there exist *no* associations of structural with fruiting variables except at a high risk. The interrelations between its structural variables are similar to population 078A. However in the present population, DNEq is in a *negative* association with DNA at a fairly low risk. That is, if the nearest alien plot is exceptionally close to the plot in question, then the other classes are (on average) unusually far from it.

In conclusion, in only *one* (004A) of the

Table 16. Correlations of quantities in population 186A. Lower number is Kendall's rank correlation coefficient ( $\tau$ ), upper number is the risk probability in a one-tailed test.

	BBS	NBB	PBB	PUF	NOF	NF	C%	DEq	DNEq	DNA	IN
BBS	•	<.0001 .67	<.0001 .69		.51 -.01	.0072 .31		.74 -.09			.12 .16
NBB		•	<.0001 .60		.49 .00	<.0001 .50	.0030 .34	.93 -.20			.19 .12
PBB			•					.52 -.01			.40 .03
PUF	.039 -.24	.094 -.18	.012 -.30	•	.43 .02	.69 -.07	.70 -.08		.77 -.11	.86 -.16	
NOF			.54 .01		•	.43 .02	.36 .04	.51 -.01	.46 .01	.40 .03	
NF			.81 .12			•	<.0001 .51	.9944 -.34	.94 -.21	.83 -.13	
C%	.79 .10	.77 .10					•	.9951 -.34	.46 .01	.22 .10	
DEq				.957 .23				•			.25 .09
DNEq	.14 -.15	.26 -.09	.74 .09					.66 .05	•	.0006 .42	
DNA	.10 -.18	.18 -.13	.52 .00					.36 -.05		•	
IN				.32 -.07	.45 -.02	.54 .01	.25 -.10		.012 -.30	.039 -.24	•

$H_0$ , no correlation; in the upper triangle,  $H_1$  dictates a positive correlation, in the lower a negative correlation. The choice of the counter-hypothesis  $H_1$  is explained in the text.

Risk probabilities below .05 or over .95 are presented in *italics*, and those below .01 or over .99 in **bold face** type.

four populations considered, the fruit set results *within* a population (ie. the differences in berry production between plots), could be explained relying at a plot level on our 'population structural' main hypothesis. In the other three populations, plot level results were indifferent or even to some extent contradictory to this hypothesis. This is not unexpected, however. Since, considering *single plots*, 'stochastic' as well as 'systematic' *variation of environment* within a population area may be overwhelmingly great in comparison to the local differences between plots caused by incompatibility classical structure. Such environmental variables, uncontrolled in the present study, could plausibly include eg. gradients in moisture, illumination or soil fertility, as well as (individually fairly constant) foraging paths (HEINRICH 1976) of bumble bees. Thus, in many populations, differences in the availa-

bility of compatible pollen for different plots, may well prove *not* to be the central causative agent for the local differences in fruit set *within* a population.

#### 4.3.8. Effect of introducing alien ramets into a population

On the basis of cross results, population 002A proved to consist of only a single incompatibility class. In its area, a plot had been prepared, into which a few arctic bramble ramets from another, nearby population had been planted (Fig. 19). Though population 002A become classified as a non-fruiting one, still some 'poor' fruits occurred, containing as a rule one or two drupelets (Tab. 9), and in addition a single 'big' berry (with at least ten drupelets). Studying their spatial distribution, one may be able to discover, whether this plot

Table 17. Correlations of quantities in population 078A. Lower number is Kendall's rank correlation coefficient ( $\tau$ ), upper number is the risk probability in a one-tailed test.

	BBS	NBB	PBB	PUF	NOF	NF	C%	DEq	DNEq	DNA	IN
BBS	•	<.0001 <i>.54</i>	<.0001 <i>.55</i>		.50 -.00	.0058 <i>.32</i>		.78 -.13			.83 -.15
NBB		•	<.0001 <i>.53</i>		.54 -.02	.0013 <i>.38</i>	.011 <i>.29</i>	.80 -.13			.83 -.15
PBB			•					.91 -.21			.78 -.12
PUF	.83 <i>.12</i>	.81 <i>.11</i>	.89 <i>.16</i>	•	.40 .03	.23 <i>.09</i>	.011 <i>.29</i>		.9970 <b>-.41</b>	.56 -.03	
NOF			.47 -.01		•	.059 <i>.20</i>	.47 .00	.40 .04	.83 -.15	.017 <i>.31</i>	
NF			.981 <i>.26</i>			•	.0021 <i>.36</i>	.64 -.06	.58 -.04	.051 <i>.24</i>	
C%	.94 <i>.19</i>		.976 <i>.25</i>				•	.70 -.09	.58 -.04	.90 -.19	
DEq				.48 -.01				•			.60 -.05
DNEq	.90 <i>.18</i>	.91 <i>.20</i>	.92 <i>.21</i>					.28 -.09	•		.60 -.04
DNA	.48 -.01	.34 -.07	.34 -.07					.91 .19		•	
IN				.42 -.04	.051 -.25	.0050 <b>-.38</b>	.27 -.10		.52 .00	.025 -.30	•

H<sub>0</sub>: no correlation; in the *upper* triangle, H<sub>1</sub> dictates a *positive* correlation, in the lower a negative correlation. The choice of the counter-hypothesis H<sub>1</sub> is explained in the text.

Risk probabilities below .05 or over .95 are presented in *italics*, and those below .01 or over .99 in **bold face** type.

with its presumably 'alien' genotypes had any local influence on the fruit set.

According to the 'environmental' explanation, differences in fruit set are caused by some external factors, irrespective of the genotypic spatial structure in the population. Such factors could be eg. environmental variation in space, or differences in the general genetic background (apart from incompatibility genotypes) of the ramets in the population. Self-incompatibility in the species might be variable in strength, depending on environment and the general genetic background. Hence, due to certain environmental conditions, it could sometimes not hold true. In any case, the *location* of the studied plots in relation to the 'alien' plot, should bear *no* relevance to the productivity of berries.

On the contrary, according to the 'popula-

*tion structural'* explanation, the 'alien' plot with its alien incompatibility genotypes, should have been decisively important concerning (even the slight) fruiting in population 002A. Having such a simple situation, the fruiting variables should show a clear *correlation* with the *distance* from the 'alien' plot. Quantities based on the number of big (NBB) or small (NSB) berries, should express a positive association with the distance in question, while the quantities based on the number of unfertilized past flowers (NUF), should show a negative association with it.

When tested, both the number (NSB) and proportion (PSB) of small berries proved to be in a *strong negative* association with the distance from the 'alien' plot (Kendall's  $\tau = -.32$ ,  $P = .0065$  for the former, and  $\tau = -.37$ ,  $P = .0016$  for the latter). The num-



Table 18. Correlations of quantities in population 102E. Lower number is Kendall's rank correlation coefficient ( $\tau$ ), upper number is the risk probability in a one-tailed test.

	NSB	PSB	PUF	NOF	NF	C%	DEq	DNEq	DNA	IN
NSB	•	.0025 .39		.32 .06	.11 .16	.49 .00	.52 -.02			.67 -.08
PSB		•					.27 .09			.77 -.14
PUF	.0035 -.39	.0010 -.45	•	.53 -.01	.66 -.07	.75 .09		.44 .02	.59 -.05	
NOF		.55 .01		•	<.0001 .49	.0003 .43	.69 -.09	.54 -.03	.16 .15	
NF		.68 .07			•	<.0001 .61	.77 -.12	.35 .05	.43 .02	
C%		.27 -.09				•	.54 -.02	.027 .29	.84 -.16	
DEq			.33 -.09				•			.31 .07
DNEq	.37 -.06	.47 -.02					.982 .31	•	.955 -.26	
DNA	.57 .02	.62 .05					.20 -.14		•	
IN			.82 .14	.25 -.11	.46 -.02	.67 .06		.77 .11	.0035 -.41	•

$H_0$ , no correlation; in the *upper* triangle,  $H_1$  dictates a *positive* correlation, in the lower a *negative* correlation. The choice of the counter-hypothesis  $H_1$  is explained in the text.

Risk probabilities below .05 or over .95 are presented in *italics*, and those below .01 or over .99 in **bold face** type.

ber of unfertilized past flowers (NUF) did not show any association with the distance from the 'alien' plot except at a very high risk ( $\tau = -.08$ ,  $P = .29$ ). The respective proportion (PUF), however, was in a *strong positive* association with the distance considered ( $\tau = .38$ ,  $P = .0011$ ).

Regarding the location of the single big berry in the population, a *randomization test* based on Fisher's randomization principle (eg. SIEGEL 1956, p.152—154) was constructed. Supposing that the distance from the 'alien' plot had no effect whatsoever, the probability can be calculated that a berry (randomly assigned into the plots) would hit *at least as*

*near* the 'alien' plot as the berry actually observed. Actually, the only big berry was found just in the *nearest* plot. Hence, the risk probability given by our randomization test will be  $P = 1/30 = .033$ . If we take care not only of plots but also of the accumulated number of flowers in them (NPF), then a still smaller risk probability will be achieved, ie.  $P = 16/696 = .023$ . Taken together, the results from our simple experiment with population 002A clearly *contradict* the 'environmental' hypothesis but support the 'population structural' hypothesis for lacking fruit set in vigorous arctic bramble populations.

## 5. Discussion

### 5.1. Breeding system

In the present study, the definition from RIEGER et al. (1968) will be adopted. A term '*breeding system*' is used to cover all those variables apart from mutation which affect the genetic relations of the gametes that fuse in sexual reproduction. According to LEWIS and JOHN (1964), two main groups of such variables may be distinguished.

1. Those variables which affect the *ability* of particular gametes to fuse or parents to mate (ie. the variables comprising the '*mating system*'), and 2. those variables which affect their *probability* within the limits set by the first. The breeding system controls the extent of outbreeding which may take various forms: exclusive or predominant outcrossing (due to eg. self-incompatibility), predominant selfing, and a mixture of selfing and outcrossing.

According to HARPER (1978), the *clonal growth* habit is usually tightly linked with strict *outbreeding* (dioecy or self-incompatibility).

This linkage is so tight that LEVIN and KERSTER (1971) utilize it in 'characterizing' a clone — as a group of organisms having a strong correlation in space, but being incapable of sexual reproduction inter se.

Previous results (TAMMISOLA and RYNNÄNEN 1970) already indicated that arctic bramble would fit well into this general picture. It displays a well developed clonal growth through rhizomes, and due to the mating system (self-incompatibility), it was regarded as an outbreeder in general. The present study confirms this picture. Not a single self-compatible clone was recorded. Self-incompatibility appears to be fairly strong and universal at least in Finnish populations of arctic bramble. Hence, arctic bramble is to be regarded as a *strictly outbreeding* species.

However, since many accessions remained untested, a possibility for rare self-fertility cannot yet be excluded. Furthermore, in intraclassic pollinations, sometimes a few seeds resulted. Since a team of several persons were involved in crossing, and each year the team consisted of mostly newcomers, errors could not be totally avoided. In most cases, therefore, such seeds could have been caused by technical inconsistencies, ie. contamination and

mistakes by man. A certain proportion of them may however represent *true* cases of an "error" by nature, where an incompatible pollen tube has in fact been able to penetrate the style and enter the ovule. Such occasional errors might be due to special environmental conditions or caused by "weakness" of an incompatibility allele, or such instability might be created by a special genetic background ie. by interactions from secondary gene loci. Occasions of a sensitivity of incompatibility reaction to modifying factors, has been reported for many plants — most often on sporophytically self-incompatible species (eg. HECHT 1958, LINSKENS 1964a, b, KWACK 1965, ASCHER and PELOQUIN 1966a, b).

### 5.2. Dynamically optimized classification

The system (called *RISTO*) constructed for an automatic guidance of classificatory experiments, consists of several steps. The overall efficiency of the system is dependent on the correct functioning of each of these steps. Crossing proved fairly difficult and time consuming. However, with the aid of control crosses and careful recording of any exceptional details, and by selecting the most experienced and skilled persons for crossing work, reasonably reliable crossing data was obtained. The extended single move algorithm for clustering into equivalence classes, proved somewhat time consuming but functioned adequately. In practice, the crossing recommendations appeared to be efficient and easy to apply. In the allocation of resources between populations, attention should also have been paid to the ecological background (flowering rhythm) of the populations.

Crossing work was 'blind' and whenever possible even 'doubly blind'. That is, the persons performing the crosses had *no information* whatsoever (except the labels) concerning the populations or the ramets sampled. During the crossing phase, even myself was totally unaware of the spatial arrangement of

the original ramets in the natural population (except the overall dimensions of the population). In spite of this, in the results, in almost each population there can be found a *spatial correlation* of ramets belonging to a common incompatibility class (Fig. 19). This correlation in many cases was so evident that tedious statistical calculations would not have been necessary. Such spatial correlations provide support to the *reliability* of the experimentation system.

As an example of such a spatial correlation, equivalence classes in population 039A will be considered. This population was chosen, since it had only two equivalence classes, and the smaller of them contained no more than three ramets out of 28 that were analyzed. Therefore, the calculation effort would not be overwhelmingly great. A simple test can be constructed applying Fisher's randomization principle (eg. SIEGEL 1956, p. 152—154). Assuming that three ramets were taken by chance from the 28 available positions in the population, the probability was studied that these three random ramets would have occurred on average as close or closer to each other than the three ramets (constituting the smaller equivalence class) did in the actual population. In the population, 69 such 'close-by' combinations occurred among the total number of  $\binom{28}{3} = 3276$  three-wise combinations. Hence, the risk probability when abandoning the hypothesis of a random spatial arrangement against a hypothesis of a positive spatial association (correlation), will be  $69/3276 = .021$ . Thus, evidence for a positive spatial correlation was obtained. In populations with more numerous or more even-numbered classes, risk probability would become essentially smaller.

In the present practical example of a study, the overall efficiency of the guidance system appeared to be fairly high, leading to at least 90 per cent saving in crossing effort. At a *theoretical* level, however, there remains quite a lot to be studied. A theoretical estimate of the *efficiency* should be derived, especially concerning the crossing recommendations. The single exchange grouping algorithm requires more study, in order to solve the actual need for still further extensions. That is, one should be fairly sure of finding the actually *global* optimum point instead of only the locally optimal ones. Could this be found more easily by introducing further extensions in the algorithm or simply by adequately varying the starting point? The possibility for a *divergence*

in the iteration, might also be worthy of investigating.

With these precautions, RISTO could be *generally applicable* in any studies concerning partitioning of elements into equivalence classes. Hence it could be tried with other data and in *other disciplines*.

### 5.3. In vigorous natural populations, lack of fruit set is generally due to uniclassic constitution

At the level of population means, the results of the present study give evidence that the number of equivalence classes (NEq) and the proportion of alien flowers (PAF) are important with regard to fruit set in natural populations of arctic bramble. One might, however, insist that other factors could be more important. In order to study this latter proposition, the following hypotheses will be formulated.  $H_0$ : among non-fruiting populations, the frequency of uniclassic ones is  $\frac{1}{2}$ . Respectively,  $H_1$ : the frequency in question is greater than  $\frac{1}{2}$ .

In the results, (Tab. 9), eight non-fruiting populations occurred, out of which seven proved to be uniclassic ones. Therefore, in favour of the hypothesis  $H_1$ , the hypothesis  $H_0$  can be abandoned at a risk  $P = 8 \cdot \left(\frac{1}{2}\right)^8 + \left(\frac{1}{2}\right)^8 = .035$ . Hence, at a considerably low risk probability one can conclude that in Finland, the lack of fruit set in vigorous arctic bramble populations in nature can be explained on the basis of uniclassic constitution *alone* in most cases. That is, in less than one half of cases, any other conceivable reasons would be needed.

This conclusion is valid however only with regard to 'ordinary' years, such as represented eg. by the years 1976—77. In theory, there occasionally occurs years with such extreme (weather) conditions that also populations with many equivalence classes may often remain fruitless.

Among non-fruiting populations, population 102E constituted an exception. Namely, in it altogether five equivalence classes were

found. In the experimental field at Viikki, its ramets produced numerous berries, and nothing exceptional was noticed.

In its original 'natural' site, however, the population was not wholly normal. Its ramets were often suffering from a 'systemic' rust fungus (*Gymnoconia peckiana* (Howe) Trott.). In arctic bramble, this rust species is rare (during the collection trips, in addition to population 102E, it was found only in population 075A). When it occurs, it severely diminishes the vitality of the ramets. Furthermore, this population was also suffering from a "clay problem". Ridges of clay had been amassed along the population. Arctic bramble is said to thrive badly on clay soils (SAASTAMOINEN 1930, RYNNÄNEN 1973). The poorest ramets were certainly not included in the sample but a more vital ramet was randomly chosen instead. It may be questioned, whether population 102E should have been at all considered vigorous as a population. Hence, the fruitless condition may have been in part due to a poor general vitality of the original population.

In this instance, lack of pollinating insects might also serve as a possible explanation. Namely, the population was situated in the centre of a large open field area. During the collection trip in 1977, no honey bees or bumble bees were found in the neighbourhood of the population. Since the population was quite small and far from other sites of arctic bramble, honey bees would most likely not be interested in it. For, honey bees are highly flower constant, and therefore their interest is usually concentrated on a good nectar and pollen source offered by some *abundantly* flowering plant species (PERCIVAL 1947, FRISCH 1959). If arctic bramble flowers are scarce in the field, they will be of no interest to honey or bumble bees, since they will concentrate on other flowers eg. white clover (*Trifolium repens* L.) (KANGASJÄRVI et al. 1988).

Another deviating population was the triploid one, 116A. Its flowers were practically entirely sterile in any kind of cross. Hence, it was impossible to demonstrate a uniclassic constitution by utilizing crosses. Therefore, to be precise, we cannot *prove* that only a single incompatibility genotype was actually present in the population. The morphology of the ramets in different plots was however strikingly similar — although the common morphological features due to the triploid consti-

tution (Fig. 3) may have caused an exaggerated impression of identity. On the other hand, since polyploidy is not at all common in arctic bramble (cf. LARSSON 1957, 1969, ZHUKOVA and TIKHONOVA 1973), it is hard to believe that several different occasions of a triploidization would have occurred in a single population. On this basis, therefore, population 116A was recorded as uniclassic — and for explaining the lack of fruit set, the uniclassic constitution alone would have sufficed.

With hindsight, various "rescue" hypotheses could of course be proposed, in order to get these results to fit also with the environmental explanation. For instance, the incompatibility classical structure of the population might be stated to be solely a *result* from the fruiting situation, instead of constituting its cause. Due to an abundant fruiting, a multitude of seeds and later on seed plants, clones and equivalence classes would arise. Respectively, a poor fruiting would result in but a few seeds and seed plants, and therefore fewer equivalence classes at the end. Abundance of fruiting would be controlled — *not* by the equivalence classical structure of the population, but by environmental factors or by a genetically determined "general" production capacity peculiar to the population.

This rescue hypothesis would imply a certain degree of self-fertility. Without any demonstration of self-fertility occurring in a (natural) population, such a hypothesis is in a severe trouble from the beginning. Secondly, the lack of fruit set in nature can not be due (alone) to any in-born low level of fruiting capacity in these populations. Since, when moved to Viikki, each fruitless population (except 116A and 089B) changed into a 'normal' level of fruiting. Further evidence against it is provided by the experiment in population 002A (see 4.3.8). In the original site in nature, a small experimental change in the local proportion of alien pollen, produced a respectively localized change in fruit set.

#### 5.4. Rich fruiting presupposes many incompatibility classes

In the results (Tab. 9), there occurred ten populations with a rich fruit set. With only a single exception, each of them contained *at least three* equivalence classes of incompatibility.

The sole possible exception was provided by population 037A with two classes in the five studied ramets (Tab. 5). The 25 ramets which remained unresolved, might have provided still an extra class. Any proof is however impossible (cf. TAMMISOLA 1986, 'Richness').

In abundantly fruiting populations, the proportion of alien flowers was generally *high*: from .42 to .86 as a rule. In population

212A, this was only .32. This population however extended far outside the studied area (Tab. 1), plausibly containing extra classes and thereby a higher proportion of alien flowers in total. In population 010A, the proportion was only .18. The same comment as above will apply. In addition, the fruit number class of this population is actually less well based than in other populations (Tab. 1).

In dioecious *Rubus chamaemorus* L., hand pollination experiments indicated that pollen availability limited seed production in female dominated habitats but not in areas with an equal sex ratio (ÅGREN et al. 1986). This result fits in nicely with our main hypothesis. For a rich production of seeds in a (self-incompatible, monoecious) species, sufficient (genetically successful) pollen must be supplied. This will be achieved when there are an adequately high proportion of (alien) pollen flowers in the population (which usually presupposes many equivalence classes).

In *poorly producing* populations, different cases occurred. In population 039A, there were only two equivalence classes; in addition, one of them constituted only a small proportion of the population. Therefore, also the proportion of alien flowers remained low (.18). Alien pollen was certainly not available outside the studied area, since no arctic bramble ramets occurred there. Also in population 030A, only two equivalence classes were found. The proportion of alien flowers was fairly high (.49) — this was based on only 9 classified ramets, however. In the map (Fig. 19), some signs can be seen that the classes might have been disjointed, with one big and the other small. If this would hold, the true proportion of alien flowers would actually be lower. Pollen flow from outside the sampled area would most likely be small, since only a few ramets were found there and the nearest ones were growing at a distance of over 20 metres. The vitality of the population had decreased, since mixed forest had already grown there and caused shading. Therefore, selecting a new population as a replacement was considered. Population 078A had remained in fairly shaded conditions under alder (*Alnus*) bushes. Although a total of six equivalence classes occurred, one of them was prominent and was situated fairly disjointed from the others. The population did not continue outside the sampled area. In population 004B, the fruit number class was a little uncertain (Tab. 1) — it might have been actually classified into richly producing populations instead. Four equivalence classes were found and (on the basis of 8 classified ramets) the proportion of alien flowers was very high (.90). In the map we can however see that the population consisted of two clearly separate patches, and the equivalence classes seem to be fairly disjointed. The average distance of a classified ramet from flowers of alien

classes is as great as 13.6 metres. In addition, the ramets were scattered in strongly competing high grasses. Hence, availability of alien pollen may actually have remained much lower than would have been expected on the basis of the high proportion of alien flowers alone.

Regarding poorly producing populations, perhaps only one case (039A) would have been caused solely by incompatibility classical structure. In the three other cases, in place of (or in addition to) population structure, *unfavourable environmental conditions* (in the first place competition by other species) may have been the primary cause for the poor fruit set.

## 5.5. Other factors affecting fruit set

### 5.5.1. Equivalence class size

Provided the genets of the population are widely separated, *large* clones with numerous fertile ramets will always have a greater proportion of endogenous ('own') pollen on their stigmas than will smaller clones. For example in a wind-pollinated species, *Carex platyphylla*, the average load of endogenous pollen increased sharply with the size of the clone — up to a clone size of about 10 'culms' (ie. reproductive spikelet complexes) (HANDEL 1985). This statement should also hold true for insect-pollinated plant species. Since, foraging by lepidopterans, flies, beetles, bees and hummingbirds has to be economic in terms of energy expenditure; most flights are from a plant to one of its *nearest neighbours* (LEVIN and KERSTER 1969). In a "realistic" simulation study (LEVIN and WILSON 1978), large patches received relatively less alien pollen than small ones.

Also with the arctic bramble, alien pollen wears off quite rapidly from a pollinating insect foraging a large, disjointed clone — at least in a cultivated arctic bramble field (Fig. 21). In natural conditions this quick "wear-off" was demonstrated in population 002A. In an experiment, a couple of "alien" pollinator plants introduced in the population (see 4.3.8. and Fig. 19) caused berry production only in their *immediate vicinity*.

In a natural population, due to the lower density of flowers (see beneath), the bees should have transferred alien pollen further than in a cultivated field. Plausibly the total number of flowers functioning as alien pollen

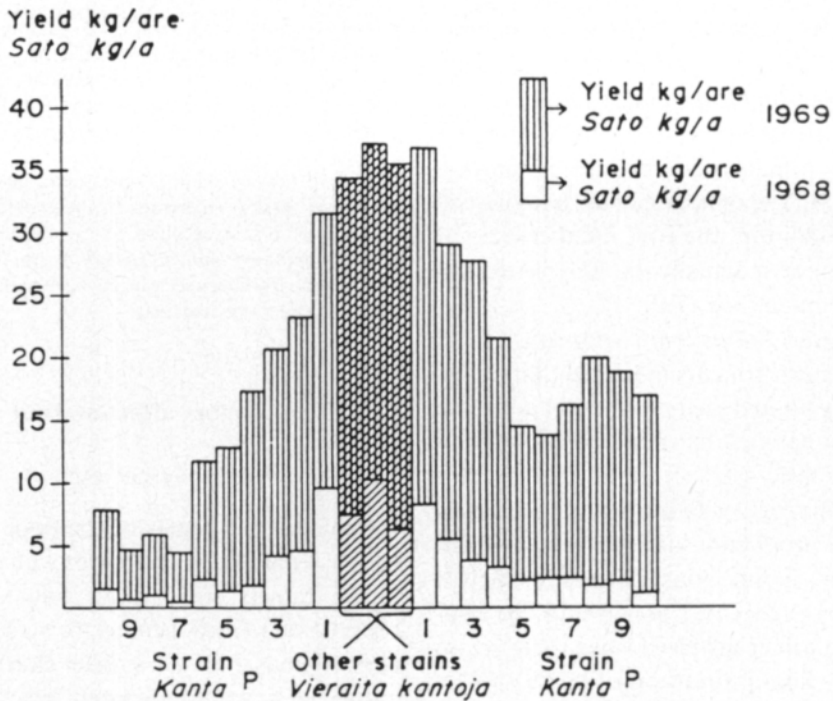


Fig. 21. Yields of an arctic bramble strain at various distances from other strains. Ten rows of strain P were planted at 1 metre intervals on each side of three centre rows of alien pollinator strains. Fruit yield was recorded in the years 1968–69 (RYYNÄNEN 1973).

donors, was however too small to demonstrate such an effect.

### 5.5.2. Density, aggregation and patch form

The denser the population, the shorter will be the flight intervals (LEVIN and KERSTER 1969). Therefore, presuming widely separated genets, the more aggregated is the distribution of ramets, the greater should be the proportion of endogenous pollen on the stigmas, as GLEAVES (1973) states. In the simulation study of LEVIN and WILSON (1978), the alien pollen influx appeared to be a function of patch form as well. Within a continuous population, elongated patches received relatively more alien pollen than square-shaped ones with an equal area.

In a cultivated arctic bramble field, the situation is not so simple, however. Preliminary results of KANGASJÄRVI et al. (1988) indicate that bees as well as bumble bees predominant-

ly keep foraging along a planted row and only rarely (in less than 4 % of the flights) will change the row. Due to this foraging behaviour, elongate patches (ie. rows) of arctic bramble appear to receive much less alien pollen than expected on the basis of LEVIN and WILSON'S (1978) model.

On the basis of this result by KANGASJÄRVI et al. (1988), one might wonder, why so many berries were produced so far from the alien strain in RYYNÄNEN'S (1973) experiment (Fig. 21). However, due to bee hives, pollinator density was then fairly high. A certain background effect may have been exerted by other arctic bramble fields which were nearby. Furthermore, the strains used by Ms Ryyänen in the experiment were at that time not genuinely pure but to a small and uncertified degree mixed with each other and their occasional seed progeny. These factors may have generated a certain "background level" of alien arctic bramble pollen in the experiment. A more profound reason for the difference may stem from differences in cultivation methods. Kangasjärvi and others used a plastic mulch technique which is commonly used for strawberry. This resulted in strong concentration of flowers into the centre of the row. While in Ryyänen's experiments, arctic brambles were growing in free rows (covered with a mulch of gravel or moss) which were almost grown together. Hence, in her field, the belts of arctic bramble

flowers were much broader and less dense, and less clearly separated from each other, which likely resulted in essentially less "row constancy" from the pollinating insects.

Since *natural populations* are rarely arranged in nearby rows — even on road banks arctic bramble usually thrives much better on one side — "row constancy" of pollinators should not have much influence on their fruit set. It is a fact that in this study there existed linear populations (eg. 004A, 039A). But since they consisted of only a single row each, the effect of row constancy cannot be separated from flower constancy of the pollinators.

For a conclusive study into the effect of density, aggregation or patch form, I consider the data of the present study to be too *robust*. Thirty studied ramets out of hundreds or even thousands present, gives far too little control of the spatial structure of equivalence classes in the population. Furthermore, utilizing information from different populations, will usually not increase the exactitude of the information. Since, the observations have been made in different places and at widely different times, and so many factors in the complicated plant-pollinator system have been left without control or remained unnoticed. Therefore, any details and minute differences will inevitably be buried under the robust mass of "change" variation remaining. In order to find out the effect of any such variable on fruit set, one should construct and start *special experiments* particularly designed for the purpose.

### 5.5.3. Fruiting zone

Concerning the possible differences in the proportion of non-fruiting populations between the fruiting zones proposed by SAASTAMOINEN (1930), *no* information whatsoever can be obtained in the present study. This is due to the stratified sampling — one just aimed at acquiring fruitless populations *in equal proportions* from each fruiting zone.

## 5.6. Vegetative or generative burst

WHITNEY (1978) has studied growth and

reproduction in *Rubus idaeus* L. Its populations arise largely from the *germination* of buried seeds triggered by environmental stimuli associated with disturbed site conditions. Hence, its dense populations are very even-aged. As conditions for vegetative growth deteriorate, *R. idaeus* gradually shifts to the production of large numbers of seed. The berry productivity is usually at its highest during the fourth year. In *R. idaeus*, seed decay rate is extremely low, with a half-life which according to WHITNEY (1978) probably approaches one hundred years. He however presented no data in favour of such a claim. The association of relatively high flux in the actively growing shoot portion of the population, coupled with a low flux rate in the seed bank fraction, is regarded as an adaptation to an *intermittently* favourable environment.

Also *arctic bramble* is well known to "appear suddenly" in dense stands at certain places suitably disturbed by man or nature (eg. by burning, grazing, slight ploughing, or clearing). It is known to be a weak competitor, with the result that rich berry production will not continue over many years, unless the site will be recurrently disturbed (eg. SAASTAMOINEN 1930, RYYNÄNEN 1973, TAMMISOLA 1981). Hence, one might reason that its populations likely will follow the same strategy as *R. idaeus*, ie. vigorous populations will appear as a result of a *generative* burst.

An alternative explanation, based on a *vegetative* burst, have also been proposed. Arctic bramble is assumed to be vegetatively present — though strongly suppressed — at the site already before the occurrence of the disturbance (SAASTAMOINEN 1930). When free space becomes available, arctic bramble would rapidly fill it up with the aid of its efficient rhizome system. ERVI et al. (1955) even postulate that arctic bramble could hide for years in the ground, being present exclusively as dormant rhizomes. *A priori*, these suggestions can be regarded as not at all unlikely, since in clonal perennials, ramets are often able to persist for a long time in the pregener-

ative states, attaining the mature state as soon as appropriate ecological niches become vacant. Even plants existing in a senile state, may due to environmental change reverse into a generative state (RABOTNOV 1978).

Utilizing the method constructed in the present study, the most central factor with regard to intensity of fruiting, could be analyzed. That is, the *equivalence classes of incompatibility* in the population (Fig. 19). Direct information concerning clones was more meagre, however. Actually, only the *minimum* number of clones present in the population was unravelled. In principle, equivalence classes and clones cannot be equated. Namely, it is quite possible that a pair of ramets are identical at the incompatibility locus but quite different at other loci, and thus belong to different clones. The spatial correlation of ramets belonging to a common incompatibility class (see 5.2.) however suggests that equivalence classes could usually be equated with clones. In any case, regarding populations with only a *single equivalence class*, it is the *most likely* explanation that all the ramets belong to a common clone, i.e. are originally derived from a single seed. This explanation is still not unique, since both an origin from two seeds and from three seeds prove to be fairly likely hypotheses.

A population with only two alleles of the incompatibility gene, would be fruitless. Hence it would have originated from a population with more alleles. It can be shown (App. L) that if we take  $n$  seeds at random from a population with three incompatibility gene alleles, the probability that all of these belong to a *common* incompatibility genotype will be less than  $1/2^{n-1}$  but at least  $1/3^{n-1}$ . The smaller of these figures will apply, provided each incompatibility genotype produces an equal amount of seeds in the population. If on the contrary almost all of the seeds in the population were produced by a single incompatibility genotype, then the greater figure would serve. This latter situation would only rarely be met in a population, since in extreme situations, gametophytic incompatibility provides a *negative feedback* mechanism between the proportion of ramets and the proportion of seeds. Thus, for two seeds the probability of identical incompatibility class will be about  $1/3$  (or at least less than  $1/2$ ), for three seeds about  $1/9$  (less than  $1/4$ ) and for four seeds about  $1/27$  (less than  $1/8$ ). Therefore, applying the principle of maximum likelihood, the likelihood value for an origin from *one seed* is about three times higher than the likelihood value for an origin from two seeds, and about nine

times higher than that for the origin from three seeds. In case of more than three incompatibility alleles represented in the original population, the one seed origin of the present population would be still more likely in relation to the alternatives. Hence, the maximum likelihood *estimate* of the origin of the population is that it is uniclonal. The risk probability when abandoning the two-seed origin is however fairly great (about  $1/3$ ).

Uniclassic populations will therefore provide a preliminary estimate concerning *clone* (genet) *size*. In a pilot study (TAMMISOLA 1981), the maximum distance between two ramets in a clone, was estimated as 38 metres. In the present study, an extra sampled ramet at the distance of 60 metres from the uniclassic population 099A, proved to belong to the common class. Therefore, the minimum size of the clone was estimated to be 80 metres. One should however notice that no attempts at searching for any records in clone size were included in the study program. Hence, possibilities for even larger equivalence class sizes still remain.

Taking 25 cm as the height of the aerial shoot, then the width/height ratio in arctic bramble reaches 400, which is one of the greatest values hitherto recorded in vascular plants (see Fig. 1, in TAMMISOLA 1986).

From experience in cultivated field conditions, under favourable circumstances, the development of growth in one direction may be half a metre per year, and sometimes even more. However, if strong competition is exerted by weeds, then the spread of arctic bramble may cease altogether. Concerning the growth rate of arctic bramble in *natural* conditions, no reliable observations have been made. Considered over a long time interval, this must however be on average less than the value recorded in cultivation. Since in nature, intermittently troubles with eg. soil quality, and also from a severe competition by other plant species will occur. Eg. the radial growth of Finnish bracken populations (*Pteridium aquilinum* (L.) Kuhn.) seems to have averaged less than 20 cm a year (OINONEN 1967). Postulating that arctic bramble will grow on average 25 cm per year in one direction, the minimum *age* of the single clone in population 099A is estimated to be 160 years. Other populations would yield a clone of at least 80



years and several of at least 30 to 40 years old. Notice that these populations were not yet deteriorating but generally still vegetatively vigorous and in a vividly generative stage. Considering these facts, although the estimation of clone ages suffered from some uncertainties, such high potential ages may be regarded as an evidence in favour of the hypothesis of a *vegetative* burst.

Further evidence for a possibility of great age in an arctic bramble clone, may be obtained by considering a hybrid between arctic bramble and stoneberry (*R. saxatilis* L.). This hybrid is quite frequently met in Northern Finland. Since it is as a rule triploid ( $2n = 21$ ), the hybrid is usually sterile in spite of abundant flowering (VAARAMA 1939, 1949, 1954). Therefore, in most places, its maintenance must be based exclusively on vegetative means. Along the river Sotajoki in 1976, in an interval of some hundreds of metres, ramets of this hybrid were found. Tested at Viikki, the hybrid set no seed and proved to spread under soil level just like arctic bramble. Experience shows that such hybrids from different districts often differed fairly clearly in morphology from each other. These ramets at Sotajoki were however similar in appearance. Although possible, an occurrence of several different hybridizations at nearby sites is to be regarded as less likely than an occurrence of a single one. The lowest portion of the bank near water would have provided the hybrid with a permanent niche for decades or centuries. Therefore, though a direct proof is still missing, it can be considered plausible that all these ramets constitute a single and very old clone.

On the basis of so uncertain age estimation, it would be unwise to present any conclusions concerning the *age distribution* of clones in a population. One should notice, however, that in many populations the clones (genets) appear to be of fairly *different size*. In part this may be due solely to change: some of the clones may have remained almost totally outside the selected study area. Nevertheless in some cases (eg. 039A), the whole population was situated within the sampling area.

Such differences in size might suggest *age differences* between the clones. That is, all clones in a population had not emerged at a same time but seed plants had arisen at quite different times. This would give evidence *against* the hypothesis of a (purely) *generative* burst. Since, according to this hypothesis, the clones should be of a similar age (as was the case in *R. idaeus* referred above). Additional explanations for the different sizes of the

clones can be proposed, however. The clones may have a different *growth rate* in vegetative spread, or have a different *competitive ability* against other rivalling species. In the present study, however, such factors were not investigated. Therefore, on the basis of the data in this study, no steadfast answer can be given concerning the possible age differences of the clones.

Nevertheless, at one stage or another during the history of these populations, germination of seeds and establishment of seed plants *has* occurred. Therefore a *mixed* model seems to be most appropriate, stating that both vegetative and generative burst are functioning — although in different populations and at different times with different relative intensities.

In principle, a *uniclassic* arctic bramble population may have originated from a distant colonization by a single seed. There is still another possibility, too. The population may have originally contained many equivalence classes, only one of which has remained alive (or non-dormant). The present study yields no information utilizable in the choice between these two alternatives. Since, the analysis of flowering ramets into incompatibility classes, will concern only the structure of the population at the *time of collection*. In order to obtain information on the *history* of the population, the seed flora buried in the soil could be investigated. In order to survive, a new population emerging as a result of a distant colonization by a single seed, will have no alternative but to apply the vegetative burst. Since, no seed bank of arctic bramble will be available. A uniclassic rudiment of an earlier polycyclic population, may however still for some time utilize also a generative burst. Namely, there is a lot of arctic bramble seeds in the seed bank, and these will perhaps retain their germinability at least for some decades (direct evidence is however missing).

### 5.7. Plant breeding in natural populations

The norms of 'purity' and 'absence of var-

iation', applied in the administrative definition of a plant variety, have lagged behind the development occurring in the theory and practical goals of plant breeding. In order to increase the resistance against diseases, and more generally to obtain a better ecological tolerance in cultivated plants, a *controlled*, deliberately generated *variation* should be introduced into the cultivated plant populations. New and efficient methods in plant breeding and multiplication have been utilized, which renders the production of homozygous individuals and vegetative multiplication of plants in a large scale possible in ever increasing number of species. Due to such methods, *line mixtures* and *clonal assortments* will certainly be taken into use for many plant species. Hence, in the administrative norms con-

cerning varieties, we should draw farther from the formal requirement of homogeneity (as has already happened even to the archaic species concept in botany). The units of a cultivated plant should be understood to be and defined as quantities at a *populational* level.

To *define* plant breeding fairly broadly: it constitutes *changing the genetic constitution of a plant population in order to optimize its yield*. Here the term 'yield' should be understood appropriately in different cases; eg. in ornamental plants it is different from cereals. So defined, plant breeding could be practised also in natural populations *in situ*. A person who is planting alien arctic bramble ramets in a uniclassic natural population — in order to call forth a fruit set — could be regarded as a plant breeder.

## 6. Summary

Arctic bramble is known to be a weak competitor. It usually thrives only during a few years in areas, where favourable growth conditions have been provided for it by a suitable disturbance in the environment, eg. by burning, timber felling, or grazing by cattle. When the more strongly competitive species and especially shadowing trees have taken over, the arctic bramble vegetation will usually grow less dense with fewer flowers, and consequently have fewer berries. Due apparently to weather conditions, the overall berry productivity of arctic bramble in Finland also varies greatly from year to year.

However, fruit set in apparently vigorous and richly flowering populations is known to vary "irregularly" between different sites as well. Of the many explanations offered over the years, none can be regarded as a satisfactory one. Usually they are based on environmental factors, eg. illumination, moisture, temperature, snow cover, frost or the foraging behaviour of pollinating insects. It has also been suggested that the abundant fruit set in the 'richly producing zones' would be due to the predominant occurrence of self-fertile arctic bramble strains in these areas. In other regions of Finland, usually solely self sterile strains would occur. In the literature, however, not a single convincing report on the occurrence of a self fertile arctic bramble clone has been found.

In the present thesis work, the aim was to study, whether *population structure* could explain the differences in fruiting abundance between vigorously flowering arctic bramble populations in nature, and especially the occurrence of totally *non-fruiting* populations among them. Two *principal hypotheses* were introduced. First, arctic bramble was sup-

posed to be homogeneously self-incompatible over all Finland. Secondly, in vigorous populations, the *most important* factor in the intensity of setting fruit was postulated to be the *availability of compatible pollen*, ie. pollen genetically capable of fertilizing the flowers. Hence, in most cases, a fruitless population would consist of only a single equivalence class of incompatibility. Thus, in a population at least two but usually several incompatibility classes would be required in order that a rich fruit set occurs.

In arctic bramble, as is commonly found in other species in Rosaceae, the self-sterility is caused by a *gametophytic self-incompatibility* system controlled by a single gene locus. Two arctic bramble clones with a common genotype with regard to the incompatibility locus, cannot fertilize one another. It was shown that this kind of an incompatibility system will serve as an *equivalence relation*. Hence, individual arctic bramble ramets can be grouped into *equivalence classes of incompatibility* (in short: incompatibility classes), so that pollination within such a class will be unsuccessful, while pollination *between* different classes will yield a (big) berry.

In the present study, a sample consisting of 22 populations from different parts of Finland was investigated. Non-fruiting, poorly fruiting and abundantly fruiting populations were included. From each of them, 30 ramets were taken at random, dug up and transported for planting in an experimental field at Viikki, Helsinki. Crosses were then made within these 22 populations under study, in order to group the sampled ramets into incompatibility classes.

Proceeding by traditional methods, by a series of (half) diallel crosses, would have

demanded an input of human labour that would have been too great to have been performed in practice. Therefore, some methodological novelties had to be introduced instead. The aim was to make only such crosses, which would likely be most *informative* in correcting the hitherto achieved classification estimate. For this purpose, a computer aided *guidance system* RISTO was constructed. The work was organized into *cycles*. In each cycle, the system analyzed the hitherto accumulated cross results and gave *recommendations* for the forthcoming most likely informative crosses.

In a population, the partition of the ramets into *incompatibility classes* was estimated on the basis of cross results, applying the principle of *maximum likelihood*. A search was made for the maximally likely classification (among the  $8.47 \cdot 10^{23}$  different possible classifications) by utilizing a *single move* algorithm constructed for the purpose. This algorithm proceeded from one grouping to another by changing only the classification of a single ramet at a time. (Its extended version however tried also a pooling of two groups, whenever an advance was no longer available via single moves).

The *number of seeds* in a cross was considered as a *stochastic variable*. Since the probability distribution of the number of seeds was *different* in crosses within than in crosses between equivalence classes, the classification of ramets could be based on this variable. To achieve a common probability distribution over different populations, a *relative* seed number was introduced. That is, recorded seed numbers were divided by the average seed number from compatible crosses in the population under consideration. The two *probability distributions* of the relative seed numbers — one for intraclass and the other for interclass pollinations — were estimated *non-parametrically* by a modified version of Parzen's kernel estimation method. In the final steps, probability distribution and incompatibility class estimates were adjusted together by

an *iteration*. Thus, a jointed estimate was achieved. Such an iteration always converged quite rapidly, at the fourth to seventh iteration round. In each case, starting even at fairly "exotic" initial values, these iterations converged to almost *identical* end results.

For *crossing recommendations*, a strategic principle called *maximum liability* — *maximum stability* was introduced. A complete proof could not be given, however, that this strategy would inevitably yield to maximally efficient crosses. Any estimate of the minimum or average saving in cross numbers or in time, could not be presented, either. Therefore, from the point of view of pure statistical theory, the method can still be regarded as a partially heuristic one. However on the basis of *practical* results in the present study, the recommendations proved quite *efficient*. Guided by RISTO, the number of required crosses remained at about *10 per cent* of the number that would have been required by applying the traditional half diallel method in a simpler, deterministic system. Hence, in the present study, no more than about 1200 (net) crosses were needed.

•

In the populations, and among the correspondents' accessions, a series of self pollinations was made in order to find out *self-compatible* arctic bramble individuals. *None* was found. In Finland, therefore, arctic bramble proved fairly uniformly self-incompatible. However the possibility for a rare occurrence of self-compatible clones cannot be excluded.

In vegetatively vigorous and richly flowering populations, the *number of equivalence classes* proved decisively important for the abundance of fruit set. In such populations, the *lack* of fruit set was almost always due to a *uniclassic* constitution: only ramets of a single incompatibility class were present. For a rich fruiting in a population, usually the presence of at least *three* equivalence classes were required.

Out of the eight *non-fruiting* populations, seven were uniclassic ones. There occurred only one population, which contained more than one incompatibility class but still produced no berries. In the experimental field this population set fruit abundantly. Hence, its fruitless condition in nature most probably had been caused by some "external" factor, eg. lack of pollinating insects. One of the populations proved to be triploid, and consequently it did not set fruit in any of the crosses. In populations setting fruit *abundantly*, usually from three to seven equivalence classes occurred. Only one population with possibly only two classes was encountered. In *poorly* producing populations, from two to six incompatibility classes were found. However, environmental factors (eg. competition by other species) often made it unfavourable for the arctic bramble. In these cases, some of the populations should perhaps not have been considered as fully vigorous any more.

The most *likely* explanation for a uniclassic

population is that it originated from a *single seed*, and the plant has spread over a large area via the rhizome. An origin from two and even from three seeds, is still a fairly likely hypothesis. Adopting the most likely explanation, the largest arctic bramble clone found in this study, was *80 metres* in diameter. Consequently, its age could be calculated to be at least 160 years. Such a long time span would give support to the hypothesis of a *vegetative burst*. This states that arctic bramble vegetation would *stand* for a long time — although suppressed and hard to notice — at the former places where it flourished. Due to its efficient root system, and following a suitable disturbance, arctic bramble would then succeed to reclaim the site in a few years with a dense population of shoots. Provided there are seeds from these favourable periods buried in the ground — even decades ago — production of *seed plants* may also on some occasions be involved in the burst.

## 7. Conclusions

1°. In vegetatively *vigorous* and abundantly *flowering* but however non-fruiting populations of arctic bramble in nature, generally only ramets of a single incompatibility genotype are found. Most likely these populations consist of only a single clone. By planting in the population shoots from an *alien* arctic bramble strain (from another site), such populations could be transformed into *richly fruiting* ones. Since the average transport of pollen in the population is not far-reaching, the alien ramets should occur fairly *evenly* in all parts of the population. In order to provide a sufficient supply of alien pollen, ie. an adequately high proportion of alien flowers in the population, then enough space for growth should be reserved around the planted shoots. By copying the structure of richly producing natural populations by planting *several* alien strains, this would help to ensure a high level of berry production. In principle, also commercially available varieties of arctic bramble or even of all-fieldberry, would serve as such pollen sources.

2°. Also in *cultivated* arctic bramble fields, instead of the common usage of only two varieties, the planting of at least *three* varieties is to be recommended. They should be carefully planted *mixed* in the same row — not in alternate rows as has been previously recommended. The greater the number of incompatibility genotypes that are present in the field, then the higher will be the proportion of alien flowers. This will ensure more reliable fertilization.

When relying on only two varieties, also the *ecological tolerance* in berry production will often remain poor. On many occasions, one or the other of the varieties is not well adapted to the local environmental conditions, and thus its growth or flowering is retarded. Typically the pests (eg. thrips) or diseases also cause harm mostly

on one or the other of the varieties. Therefore, in spite of one variety thriving well, the berry yield will collapse — unless a third variety has been provided.

3°. In general in wild plant species, the first step in a breeding program should be the founding of a *large genetic collection*. This holds true also for the arctic bramble. Such a collection is needed in particular, since

a) the techniques for cultivation are still in an *unstable* state (eg. cultivation in free rows, or covered with a plastic mulch, or even in hanging sacks), with the result that sudden and great changes in cultivation practices will plausibly happen, and

b) so called *primitive* features (eg. self-sterility) unfavourable in cultivation still occur, while in species cultivated already for long times (eg. bramble), such features have become eliminated, due to unconscious selection, and

c) for a certain time, in these species, genetic variation is still *available* to be collected in nature (not to forget the seed banks in the soil).

Such a collection would be needed for a *comparison* of natural strains, with the purpose of finding out those strains that would be best suited for cultivation and future plant breeding. Namely, selecting strains directly from *natural* conditions, would be quite *un-efficient*, since

a) environmental conditions in nature will *differ* quite a lot from the conditions encountered in cultivation (and with regard to many characters and many populations, genotypes x environments interaction is known to be considerably great), and

b) in a natural population, the *heritability* of a character will usually remain quite *low*, due to the great variability of the environmen-

tal conditions from one individual shoot to another.

According to the present study, only few genotypes are usually found in a natural arctic bramble population. Therefore, when performing such a collection work, resources should be allocated so as to acquire samples from as *many populations* as possible, while a few shoots from each one would be enough.

4°. In Finland, arctic bramble proved to be a self-sterile species. If it could be transformed into a *self-fertile* species instead (eg. by utilizing polyploidy, translocations together with a hidden aneuploidy, or gene mutations), a strong *inbreeding depression* would therefore result. When continuing plant breeding with such a self-fertile novice, one should straight from the beginning either

a) take care of *preventing inbreeding* during the breeding work (although the commercial yield would be produced solely via inbreeding within a single variety), or

b) by utilizing modern techniques, produce huge numbers of *totally homozygous* lines to find out the few vital ones. Proceeding in the latter way, the deleterious (lethal or subvital) alleles could be from the very beginning *eliminated* from the breeding populations (as has happened with old self-fertile species in the course of evolution).

5°. *R. arcticus* is known to sometimes grow fairly rapidly via the rhizome. After a favourable environmental disturbance, it has but a few years time to spread, since it is a weak competitor. It seems to represent the so-called *guerilla* type ie. an intermingling, exploring type of clonal growth defined by Clegg (HARPER 1978). Such a rhizomatous habit is typical also for a large number of understory herbs of the temperate forests of North America, and it may be considered an adaptation to the *paucity* and *uneven distribution* of resources in the forest floor (SCHELLNER et

al. 1982). In addition, it provides the clonal patches with *quicker mixing* together, resulting in better seed set earlier in their development (HANDEL 1985).

In a *cultivation* environment, arctic bramble is provided by man with sufficient nutrients as well as with close-by pollen donors. Therefore, in these circumstances, the relatively sparse and *far-reaching* growth habit evolved during evolution, would be considered no more profitable but deleterious. Especially if arctic bramble is to be cultivated in hanging sacks or limited by a plastic mulch as with strawberry, a waste of plant resources in the production of rhizomes should be *diminished* by plant breeding efforts.

6°. On the basis of the present study, certain efforts of *future investigations* are called for. Before starting a program to construct an artificial self-fertility, more strains from nature should be screened in the hope that *self-fertility* might occur. Direct studies on *clone size* in natural populations should be started, preferably utilizing the most modern, adequately powerful methods, eg. DNA fingerprinting (JEFFREYS et al. 1985a, b). The *history* of present (especially uniclassic) populations should be investigated, by studying the seed bank and long-term germinability of arctic bramble seeds. The total *number* of incompatibility alleles in Finland should be estimated. The possible *regional pattern* of incompatibility alleles or genotypes also deserves investigation. The new methodology introduced in the present study, calls for an estimate of the *degree of saving* to be achieved. In addition, more experiments with different data should be made in order to assess its overall reliability. In principle, this is a *general* procedure for an optimal guidance of any equivalence classificatory experiments. Hence, this method should be tested also within *other disciplines* and on different problems.

**8. Acknowledgements.** I wish to express my gratitude to the hundreds of Finns participating in the arctic bramble survey of 1976—77. Without their contributions, a meaningful choice of populations for the study would not have been possible. With regard to the success of the study, an inestimable contribution was provided by my extraordinarily careful field assistants, and Anne Nurmi, Marjo Riikonen, Pirjo Saarinen and Riisa Virtanen all deserve praise. Ms Sylvi Lehtonen has drawn most of the figures. In the practical realization and execution of the guidance system on a computer, invaluable support was provided by Martin Floman and Leena Sadeniemi of the University Computing Centre. Critical comments concerning the study were gratefully received from M.Sc. (Agr.) Peter Joy and corrections of language from M.Sc. Robert Harper.

This study or its forerunning equivalent has been in part

financed by grants from The Foundation for Research of Natural Resources in Finland (1974), The Finnish Cultural Foundation (1976 and 1978), The Emil Aaltonen Foundation (1977) and The Juho Vainio Foundation (1979).

I wish to thank my little daughters Outi and Seila for preventing the licentiate syndrome, and my wife Heljä who has avoided pursuing a career, hence making it comfortable for me to play at science. The opponent, Prof. Bengt Olle Bengtsson has been patiently waiting, instrument at the ready, for the first bars. The instructor, Prof. P. M. A. Tigerstedt, is commended for a tenacious though discreet prompting, and Doc. Veikko Koski and Ass. Prof. Y. L. A. Mäkinen for their continued persistence, thanks to which the originally boundless heap of imaginary hypotheses and factual dough has been finally baked into a presumably scientific loaf.



## 9. References

- ÅGREN, J., ELMQVIST, T. & TUNLID, A. 1986. Pollination by deceit, floral sex ratios and seed set in dioecious *Rubus chamaemorus*. *Oecologia* (Berl.) 70: 332—338.
- ANON. 1976. [An inquiry about bee keepers in Finland.] Market Study Institute of the Pellervo Society.
- ANON. 1980. Kulotustoimikunnan mietintö. [Report of a committee on controlled burning in forestry]. (In Finnish). Komiteamietintö 1980: 1. Valtion Painatuskeskus, Helsinki, 109 p.
- ARASU, N.T. 1968. Self-incompatibility in angiosperms: a review. *Genetica* 39: 1—24.
- ARKIN, H. & COLTON, R.R. 1963. Tables for Statisticians, 2<sup>nd</sup> ed. Barnes & Noble, New York, 168 p.
- ASCHER, P.D. & PELOQUIN, S.J. 1966 a. Effect of floral aging on the growth of compatible and incompatible pollen tubes in *Lilium longiflorum*. *Amer. J. Bot.* 53: 99—102.
- & PELOQUIN, S.J. 1966 b. Influence of temperature on incompatible and compatible pollen tube growth in *Lilium longiflorum*. *Can. J. Genet. Cytol.* 8: 661—664.
- BEUTLER, R. 1954. Über die Flugweite der Bienen. *Z. Vergl. Physiol.* 36: 226—298.
- BONEVA, L.I., KENDALL, D.G. & STEFANOV, I. 1971. Spline transformations: Three new diagnostic aids for the data analyst (with Discussion). *J. R. Statist. Soc. B* 33: 1—72.
- CONOVER, W.J. 1971. Practical Nonparametric Statistics, 2<sup>nd</sup> ed. Wiley.
- CZAPIK, R. 1981. Embryology of *Rubus saxatilis*. *Acta Biol. Cracov Ser. Bot.* 23: 7—14.
- DAVID, F.N., KENDALL, M.G. & BARTON, D.E. 1966. Symmetric function and allied tables. Cambridge Univ. Press, Cambridge.
- EAST, E.M. 1940. The distribution of self-sterility in the flowering plants. *Proc. Amer. Phil. Soc.* 82: 449—518.
- EDWARDS, A.W.F. 1972. Likelihood. Cambridge Univ. Press, Cambridge.
- ELVELAND, J. 1983. Slätter och träda på en älvstrandäng i Tornedalen. Summary: The vegetation on a mown and on an abandoned rivershore meadow at Torne River, N Sweden. *Sv. Bot. Tidskr.* 77: 225—234.
- ERVI, L.O., HANIOJA, P. & KIVINEN, E. 1955. Mesimarjan (*Rubus arcticus* L.) marjontaa koskevia tutkimuksia. [Studies concerning the fruit set in arctic bramble (*Rubus arcticus* L.)]. (In Finnish). *Acta Agr. Fenn.* 83: 93—112.
- FISHER, R.A. & YATES, F. 1957. Statistical Tables for Biological, Agricultural and Medical Research, 5<sup>th</sup> ed. Oliver and Boyd, Edinburgh, 13 p.
- FLINN, M.A. & WEIN, R.W. 1977. Depth of underground plant organs and theoretical survival during fire. *Can. J. Bot.* 55: 2550—2554.
- FOCKE, W.O. 1910. Species Ruborum. Monographiae generis Rubi Prodrromus. Pars I. Bibliotheca Botanica 72: 1—120.
- FOWLER, C.W. & JANICK, J. 1972. Wide Crosses and Pollen Contamination in Strawberry. *HortScience* 7(6): 566—567.
- FREE, J.B. & WILLIAMS, I.H. 1972. The transport of pollen on the body hairs of honeybees (*Apis mellifera* L.) and bumblebees (*Bombus* spp. L.). *J. Appl. Ecol.* 9: 609—615.
- FRIEDMAN, H.P. & RUBIN, J. 1967. On Some Invariant Criteria for Grouping Data. *J. Amer. Stat. Assoc.* 62: 1159—1178.
- FRISCH, K. von. 1959. Aus dem Leben der Bienen. 6. ed. Berlin. 179 p.
- FRYXELL, P.A. 1957. Mode of reproduction of higher plants. *Bot. Rev.* 23: 135—233.
- GLEAVES, J.T. 1973. Gene flow mediated by wind-borne pollen. *Heredity* 31: 355—366.
- GOOD, I.J. 1977. The Botryology of Botryology. In Classification and Clustering (Ed. J. Van Ryzin), p. 73—94. Academic Press, New York, 475 p.
- HANDEL, S.N. 1985. The intrusion of clonal growth patterns on plant breeding systems. *Amer. Nat.* 125: 367—384.
- HÄRDH, H. 1976. Puutarhanviljelyä. [Horticulture; an interview by U. Iivarinen]. (In Finnish). *Maaailma ja Me* 1976 (11): 166—169.
- HARPER, J.L. 1977. Population Biology of Plants. Academic Press, London, 892 p.
- 1978. The demography of plants with clonal growth. In Structure and functioning of plant populations (Eds. A.H.J. Freyden & J.W. Woldendorp), p.27—48. North Holland Publ. Co., Amsterdam, 323 p.
- HECHT, A. 1958. Partial restoration of fertility in a sterile *Oenothera* mutant following infection by a parasitic moth larva. *Proc. 10th Int. Congr. Genet.* 2: 118.
- HEINRICH, B. 1976. The foraging specializations of individual bumblebees. *Ecol. Monogr.* 46: 105—128.
- 1979. "Majoring" and "Minoring" by foraging bumblebees, *Bombus vagans*: An experimental analysis. *Ecology* 60: 245—255.
- HIIRSALMI, H. 1971. Mesimarjan marjontaan vaikuttav-

- ta tekijöistä. [Factors affecting fruit set in arctic bramble]. (In Finnish). Puutarha 74: 626—627.
- , SÄKÖ, J. & JUNNILA, S. 1986. Jalomaarain — uusi marjakasvi. [The new arctic raspberry — a new cultivated berry plant]. (In Finnish). Puutarha 89: 340—343.
- HIPPA, H. & KOPONEN, S. 1976. Preliminary studies on flower visitors to and potential pollinators of the cloud-berry (*Rubus chamaemorus* L.) in subarctic Lapland. Ann. Agric. Fenn. 15: 56—65.
- HULTÉN, E. 1968. Flora of Alaska and Neighboring Territories. A manual of the vascular plants. Stanford Univ. Press, Stanford, 1008 p.
- 1971. The circumpolar plants. II. Dicotyledons. Kungl. Sv. Vet. Akad. Handb. 4. Ser. 13(1): 164—165.
- JEFFREYS, A.J., WILSON, V. & THEIN, S.L. 1985 a. Hyper-variable 'minisatellite' regions in human DNA. Nature 314: 67—73.
- , WILSON, V. & THEIN, S.L. 1985 b. Individual-specific 'fingerprints' of human DNA. Nature 316: 76—79.
- KALLIO, H. 1975 a. Chemical constituents of the volatile aroma compounds in *Rubus arcticus* L. subsp. *stellatus* (Sm.)Boivin, with reference to *Rubus arcticus* L. subsp. *arcticus*. Rep. Kevo Subarctic Res. Stat. 12: 60—65.
- 1975 b. Identification of volatile aroma compounds in arctic bramble, *Rubus arcticus* L., and their development during the ripening of the berry, with special reference to *Rubus stellatus* Sm. Diss., 114 p. Univ. of Turku (Finland).
- 1976. Identification of steam-distilled aroma compounds in the press juice of arctic bramble, *Rubus arcticus* L. J. Food Sci. 41: 555—562.
- & HONKANEN, E. 1975. An important major aroma compound of arctic bramble, *Rubus arcticus* L. Proc. 4<sup>th</sup> Int. Congr. Food Sci. and Technol.
- , LAINE, M. & HUOPALAHTI, R. 1980. Aroma of the berries of the hybrid *Rubus stellatus* x *Rubus arcticus*. Ann. Univ. Turku. A II 63: 17—22.
- KANGASJÄRVI, J., OKSANEN, J. & KÄRENlampi, L. 1988. Karttulan kunnan mesimarjaprojekti. Raportti vuoden 1987 tutkimuksista. [Arctic bramble project in Karttula commune. A report of the studies during 1987]. (In Finnish), 35 p. Univ. Kuopio.
- KEEP, E. 1972. Variability in the wild raspberry. New Phytol. 71: 915—924.
- KENDALL, D.A. & SOLOMON, M.E. 1973. Quantities of pollen on the bodies of insects visiting apple blossom. J. Appl. Ecol. 10: 627—634.
- KNIGHT, R.L. & KEEP, E. 1962. *Rubus* L. II. Floral biology and seed formation. In Handbuch der Pflanzenzüchtung (Eds H.Kappert and W.Rudolf), Vol. 6, p. 483—487. Paul Parey, Berlin.
- KORHONEN, P. 1979. A stepwise procedure for multivariate clustering. Diss. Comp. Centre Univ. Helsinki Res. Rep. 7: 1—77.
- KOTILAINEN, M.J. 1949. Luonnonkasvien hallavauriois-
- ta. [Frost damages in wild plants]. (In Finnish). Suom. Tiedeakat. esitelmät ja pöytäk. 1948: 137—153.
- KOTIMÄKI, M. & HIIRSAIEMI, H. 1979. Cytogenetic studies on *Rubus arcticus*, *Rubus stellatus* and their hybrids. Hereditas 91: 83—89.
- KWACK, B.H. 1965. Styler culture of pollen and physiological studies of self-incompatibility in *Oenothera organensis*. Physiol. Plant. 18: 297—305.
- LARNA, R. 1982. Mesimarjan (*Rubus arcticus* L.) muuntelusta Suomessa. [Variation of arctic bramble (*Rubus arcticus* L.) in Finland]. (In Finnish). M.Sc. Thesis, Dept of Botany, Univ. of Helsinki, 74 pp.
- LARSSON, E.G.K. 1955. Odlingsförsök med äkerbär (*Rubus arcticus*). Summary: Cultivation trials with the Arctic raspberry. Frukt i år (Sver. Pomol. För. Årsskrift) 1955: 183—195.
- 1957. Tetraploid *Rubus arcticus*, framställd genom colchicin-behandling. Summary: Colchicine-induced tetraploidy in *Rubus arcticus* L. (the Arctic raspberry). Bot. Not. 110: 152—159.
- 1968. Betydelsen av vildväxande nordliga *Rubus*-arter för framställning av nya bärsorter. Summary: The Importance of Wildgrowing Northern *Rubi* (Brambles) for Breeding New Berries. In Svensk Naturvetenskap 1968 (Ed. B. Afzelius), 4 p. Statens Nat. Vet. Forskn. råd, Stockholm.
- 1969. Experimental taxonomy as a base for breeding in Northern *Rubi*. Hereditas 63: 283—351.
- 1980 a. *Rubus arcticus* L. subsp. *x stellarcticus* subsp. nov. Bot. Not. 133: 227—228.
- 1980 b. *Rubus arcticus* L. subsp. *x stellarcticus*, a new arctic raspberry. Acta Horticulturae 112: 143—144.
- 1985. All-fieldberry. 4 p. FAKTA/Horticulture No 123 (1984, 1985 revised). Swed. Univ. of Agric. Sci., Alnarp.
- LEONARD, T. 1978. Density Estimation, Stochastic processes and Prior Information. J. R. Statist. Soc. B 40(2): 113—146.
- LEVIN, D.A. & KERSTER, H.W. 1969. The dependence of bee-mediated pollen and gene dispersal upon plant density. Evolution 23: 560—571.
- & KERSTER, H.W. 1971. Neighborhood structure in plants under diverse reproductive methods. Amer. Nat. 105: 345—354.
- & WILSON, J.B. 1978. The genetic implications of ecological adaptations in plants. In Structure and functioning of plant populations (Eds A.H.J. Freyden & J.W. Woldendorp), p.75—100. North Holland Publ. Co., Amsterdam, 323 p.
- LEWIS, K.R. & JOHN, B. 1964. The Matter of Mendelian Heredity. Churchill, London. (Ref. Rieger et al. 1968).
- LINDGREN, B.W. 1976. Statistical Theory, 3<sup>rd</sup> ed. Macmillan, New York, 614 p.
- LINKOLA, K. 1916. Studien über den Einfluss der Kultur auf die Flora in den Gegenden nördlich von Ladogasee. I. Acta Soc. F. Fl. Fenn. 45(1).

- LINNAEUS, C. 1762. Åkerbärs plantering. [Cultivation of arctic raspberry]. (In Swedish). Kongl. Vetensk. Acad. Handl. 23: 192—197.
- LINSKENS, H.F. 1964 a. Pollen physiology. Ann. Rev. Plant Physiol. 15: 255—270.
- 1964 b. The influence of castration on pollen tube growth after self pollination. In Pollen Physiology and Fertilization (Ed. H.F. Linskens), North-Holland Publ. Co, Amsterdam, p. 230—236.
- LIPSCHUTZ, S. 1964. Set Theory. Schaum, New York, 233 p.
- MÄKELÄ, K. 1984. Mesimarjan ja lakan tauteja. [Diseases in arctic bramble and cloudberry]. (In Finnish). Puutarha 87(12): 738.
- MULLIGAN, G.A. & MUNRO, D.B. 1984. Chromosome numbers and sexual compatibility in North America of *Rorippa sylvestris* (Cruciferae). Can. J. Bot. 62: 575—580.
- NOUSIAINEN, H., TERÄS, I. & VIRAMO, J. 1978. Mustikka ja puolukka — hyönteispölytteiset metsämarjamme. [Bilberry and cowberry — our entomophilous wood berries]. (In Finnish). Suomen Luonto 1978 (2): 91—94.
- NYBOM, H. 1985. Active self-pollination in blackberries (*Rubus* subgenus *Rubus*, Rosaceae). Nord. J. Bot. 5: 521—526.
- 1986. Active self-pollination and pollen stainability in some *Rubus* cultivars. J. Hort. Sci. 61: 49—56.
- OINONEN, E. 1967. Sporal regeneration of bracken (*Pteridium aquilinum* (L.) Kuhn.) in Finland in the light of the dimensions and the age of its clones. Acta Forest. Fenn. 83(1): 1—96.
- OSTER, G. & HEINRICH, B. 1976. Why do bumblebees major? A mathematical model. Ecol. Monogr. 46: 129—133.
- PAHKINEN, E. 1981. The method of support as statistical inference model for instant sample. Diss. Jyväskylän Stud. Comp. Sci. Econ. Stat. 4: 1—99.
- PARZEN, E. 1962. On Estimation of a Probability Density Function and Mode. Ann. Math. Statist. 33: 1065—1076.
- PATTERSON, T.N.L. 1968 a. The Optimum Addition of Points to Quadrature Formulae. Math. Comp. 22: 847—856.
- 1968 b. On Some Gauss and Lobatto Based Integration Formulae. Math. Comp. 22: 877—881.
- PEKKARINEN, A. & TERÄS, I. 1977. Suomen kimalaisista ja loiskimalaisista. Summary: On Finnish bumblebees and cuckoo bumblebees (Apidae). Luonnon Tutkija 81: 1—24.
- , TERÄS, I., VIRAMO, J. & PAATELA, J. 1981. Distribution of bumblebees (Hymenoptera, Apidae: *Bombus* and *Psithyrus*) in eastern Fennoscandia. Not. Ent. 61: 71—89.
- PERCIVAL, M. 1947. Pollen collection by *Apis mellifera*. New Phytologist 46: 142—173.
- POPPIUS, B.R. 1903. Blombiologiska iakttagelser. [Observations in flowering biology]. (In Swedish). Acta Soc. F. Fl. Fenn. 25.
- PYYSSALO, T., SUIHKO, M. & HONKANEN, E. 1977. Odour thresholds of the major volatiles identified in cloudberry (*Rubus chamaemorus* L.) and arctic bramble (*Rubus arcticus* L.). Lebensm.-Wiss. Technol. 10: 36—39.
- RABOTNOV, T.A. 1978. On coenopopulations of plants reproducing by seeds. In Structure and functioning of plant populations (Eds A.H.J. Freyden & J.W. Woldendorp), p.1—26. North Holland Publ. Co., Amsterdam, 323 p.
- REDALEN, G. 1976. Pollination and fruit set in raspberries. Acta Hort. 60: 169—175.
- RENYI, A. 1970. Probability Theory. North-Holland Publ. Co, Amsterdam.
- RIBBANDS, C.R. 1951. The flight range of the honeybee. J. Anim. Ecol. 20: 302—310.
- RIEGER, R., MICHAELIS, A. & GREEN, M.M. 1968. A Glossary of Genetics and Cytogenetics. Springer, Berlin, 506 p.
- RIORDAN, J. 1958. An Introduction to Combinatorial Analysis. Wiley & Sons, New York.
- RUOKOLA, A.-L. 1981. Fungus diseases of cultivated arctic bramble (*Rubus arcticus* L.) in Finland. J. Agric. Sci. Finl. 53: 83—89.
- RYYNÄNEN, A. 1971. Mesimarjan viljely. [Cultivation of arctic bramble]. (In Finnish). Kansallisosake-Pankin kuukausikats. 1971(4): 1—11.
- 1973. *Rubus arcticus* L. and its cultivation. Ann. Agric. Fenn. 12: 1—76.
- & DALMAN, P. 1983. A new variety of arctic bramble 'Pima'. Ann. Agric. Fenn. 22: 1—7.
- SAASTAMOINEN, S. 1930. Mesimarja (*Rubus arcticus* L.) Suomessa. Referat: Die nordische Himbeere (*Rubus arcticus* L.) in Finnland. Ann. Zool. Bot. Soc. Fenn. Vanamo 13: 356—414.
- SALMINEN, M. 1948. Sisä-Suomen mahdollisuuksista puutarhatuotantoon. [Possibilities for horticulture in Central Finland]. (In Finnish). Maatalous ja koetointa 3: 68—91.
- SHELLNER, R.A., NEWELL, S.J. & SOLBRIG, O.T. 1982. Studies on the population biology of the genus *Viola*. IV. Spatial pattern of ramets and seedlings in three stoloniferous species. J. Ecol. 70: 273—290.
- SCHWANITZ, F. 1967. Die Evolution der Kulturpflanzen. Bayerisches Landwirtschaftsverlag, München. 463 p.
- SIEGEL, S. 1956. Nonparametric Statistics for the Behavioral Sciences. Mc Graw-Hill, New York. 312 p.
- SILÉN, F. 1906. Blombiologiska iakttagelser i Kittilä, Lappmark. [Observations on flowering biology in Kittilä, Lappland]. (In Swedish). Meddel. Soc. F. Fl. Fenn. 31.
- SIMMONDS, N.W. 1979. Principles of crop improvement. Longman, London. 408 p.
- SOLANTIE, R. 1976. Järvien vaikutus lämpötilan mesoskaala-analyysiin Suomessa. Abstract: The in-

- fluence of lakes on meso-scale analysis of temperature in Finland. *Ilmatiet. Lait. Tied.* 30: 1—72.
- STEINHAUSEN, D. & LANGER, K. 1977. Clusteranalyse. (In German). Walter de Gruyter, Berlin, 206 p.
- TAMM, C.O. 1972. Survival and flowering of some perennial herbs. II. The behaviour of some orchids on permanent plots. *Oikos* 23: 23—28.
- TAMMISOLA, J. 1981. Principles of genetic sampling and conservation in perennial species. A case study in arctic bramble (*Rubus arcticus* L.). In *Seed Regeneration in Cross-pollinated Species* (Eds E. Porceddu & G. Jenkins), p.165—189. A.A.Balkema, Rotterdam, 293 p.
- 1986. Populations in clonal plants. *J. Agric. Sci. Finl.* 58: 239—276.
- & RYNNÄNEN, A. 1970. Incompatibility in *Rubus arcticus* L. *Hereditas* 66: 269—278.
- TERÄS, I. 1976. Flower visits of bumblebees, *Bombus* Latr. (Hymenoptera, Apidae), during one summer. *Ann. Zool. Fenn.* 13: 200—232.
- 1983. Estimation of bumblebee densities (*Bombus*: Hymenoptera, Apidae). *Acta Ent. Fenn.* 42: 103—113.
- 1985 a. Flower visits of bumblebees (*Bombus*: Hymenoptera, Apidae) during one day in northeastern Finland. *Not. Ent.* 65: 129—135.
- 1985 b. Food plants and flower visits of bumblebees (*Bombus*: Hymenoptera, Apidae) in southern Finland. *Acta Zool. Fenn.* 179: 1—120.
- VAARAMA, A. 1939. Cytological studies on some Finnish species and hybrids of the genus *Rubus* L. *J. Agric. Sci. Finl.* 11: 72—85.
- 1949. Cytogenetic studies on two *Rubus arcticus* hybrids. *J. Agric. Sci. Finl.* 20: 67—79.
- 1951. Om artkorsningsföreläring inom släktet *Rubus*. [Breeding via art crosses in the genus *Rubus*]. (In Swedish). *Nord. Jordbr.forskn.* 33: 412—417.
- 1954. Chromosome numbers of some species and hybrids of the genus *Rubus*. *Arch. Soc. Zool. Bot. Fenn. Vanamo* 8: 192—195.
- 1965. *Rubus* L. — Vatukan suku. [Bramble species]. (In Finnish). In *Suuri Kasvikirja* (Ed. J. Jalas), Vol. 2, p. 748—766. Otava, Helsinki.
- VARDJA, T. & KARIS, H. 1985. Some materials on the powdery mildew flora of Kamchatka oblast (Russian SFSR, USSR). (In Russian with Engl. summary). *Eesti NSV Tead. Akad. Toim. Biol.* 34(1): 16—22.
- VARIS, A.-L. 1981. Mehiläistalous Suomessa. [Bee culture in Finland]. (In Finnish). *Mehiläishoitaja* 15: 94—95.
- WAHBA, G. 1977. Optimal Smoothing of Density Estimates. In *Classification and Clustering* (Ed. J. Van Ryzin), p. 423—458. Academic Press, New York, 475 p.
- WALLEN, B. 1986. Above and below ground dry mass of the three main vascular plants on hummocks on a subarctic peat bog. *Oikos* 46: 51—56.
- WARMING, E. 1886. Om nogler arktiska sorters biologi. [Biology of some arctic species]. (In Swedish). *Bih. t. Kungl. Svenska Vetensk.akad. handb.* 12 Afd. III.
- WHITNEY, G.G. 1978. A Demographic Analysis of *Rubus idaeus* L. and *Rubus pubescens* Raf. The Reproductive Traits and Population Dynamics of Two Temporally Isolated Members of the Genus *Rubus*. Diss., 139 p. Yale University.
- WHITTLE, P. 1958. On the smoothing of probability density functions. *J. R. Statist. Soc. B* 20: 334—343.
- ZELLER, O. 1964. Entwicklungsmorphologische Studien an Blütenknospen von *Rubus arcticus* L. und *Rubus chamaemorus* L. in Finnland. (In German). In *Beitr. Phytol.* (Ed. K.Kreeb), p. 16—32. Eugen Ulmer, Stuttgart.
- ZHUKOVA, P.G. & TIKHONOVA, A.D. 1973. [Chromosome Numbers of some Plant Species of Chukotski Peninsula. Part 2]. (In Russian). *Bot. Zhurnal* 58(3): 395—402.

## 10. Selostus

### Inkompatibiliteetti luokat ja marjonta mesimarjan suomalaisissa luonnon- esiintymissä

Mesimarja tiedetään *heikoksi kilpailijaksi*. Se kukoistaa yleensä muutaman vuoden kerrallaan alueilla joissa sopiva ympäristön häiriö — esim. hakkuu, kaskeaminen, laiduntaminen — on varustanut sille suotuisat kasvuolot. Kun voimakkaammat kasvilajit ja erityisesti varjostava puusto ehtivät kilpailuun mukaan, mesimarjaesiintymän kunto huononee: kasvusto harvenee, kukkiminen ja marjonta vähenee.

Mesimarjan *vuotuinen* satotaso maassamme vaihtelee suuresti, ilmeisesti lähinnä säätekijöiden mukaan. Pitemmällä aikajaksolla tarkasteltuna maamme luontaiset *mesimarjasadot* ovat jatkuvasti *vähentyneet*. Pääsyyinä tähän lienee mesimarjalle sopivien *kasvupaikkojen häviäminen*. Nykyaikainen tehoamaalous ei enää juuri jätä tälle kaskikulttuurin seuralaislajille elintilaa. Säännöllinen ja syvälle ulottuva maanmuokkaus hävittää mesimarjan juurakot. Salaoituksen myötä avo-ojat ovat hävinneet pelloilta. Niitytkin ovat katoamassa. Voimakkaasta lannoituksesta hyötyvät mesimarjan kilpailijat, rikkakasvit. Laji on myös kovin arka rikkakasvimyrkyille.

Mesimarjan marjonta *vaihtelee* kuitenkin näennäisen *epäsäännöllisesti* myös hyväkuntoisesta *esiintymästä toiseen*. Kunnollista selitystä tällaiselle alueelliselle ja paikalliselle vaihtelulle ei ole löydetty, vaikka vuosikymmenien aikana onkin ehditty tarjota lukuisia ratkaisuehdotuksia. Yleensä nämä pohjautuvat ulkoisiin tekijöihin: valaistus, kosteus, lämpötila, lumipeite, pölyttäjät, halla jne. On myös arveltu, että parhaiten marjovilla alueilla valitsisivat itsefertiilit ja siksi runsastuottoiset mesimarjakannat, kun taas muualla maassa esiintyisi lähinnä vain itsesteriilejä kantoja. Kirjallisuudesta ei kuitenkaan löydy riittävää näyttöä yhdestäkään itsefertiilistä mesimarjakannasta.

Tässä väitöskirjatyössä ryhdyttiin selvittämään, voisiko *populaation rakenne* selittää marjontaerot ja erityisesti *marjomattomuuden* hyväkuntoisissa, ts. voimakas- kasvuissa ja runsaasti kukkivissa mesimarjan luonnon- esiintymissä. Marjoja ei nimittäin pitäisi ollenkaan muodostua, jos populaatiossa olisi vain *yhtä inkompatibiliteetti luokkaa*. Runsaan marjonnan taas ajateltiin edellyttävän tavallisesti useaa inkompatibiliteetti luokkaa.

Mesimarjalla itsesteriilisuuden aiheuttaa ruusukasveilla yleinen ns. gametofyyttinen itse-inkompatibiliteetti. Kaksi

yksilöä, joilla on inkompatibiliteettigeenin s osalta täsmälleen sama genotyyppi, eivät voi hedelmöittää toisiaan. Mesimarjakannat voidaan siten ryhmitellä *vastaavuusluokiksi* (inkompatibiliteetti luokat), jolloin vain luokkien *välinen* pölytys tuottaa (kunnollisen) marjan, ei luokan sisäinen pölytys.

Työssä tutkittiin 22 eri puolilta Suomea otostettua marjomatonta, heikosti tai runsaasti marjovaa mesimarjapopulaatiota. Kustakin arvottiin 30 näyteyksilöä, jotka istutettiin koekentälle Viikkiin. Populaatioiden sisäisten risteytysten perusteella näyteyksilöt sitten ryhmiteltiin inkompatibiliteetti luokiksi.

Itsepölytysten avulla em. populaatioista ja lisäksi kymmenistä yksittäisnäytteistä eri puolilta Suomea etsittiin *itsefertilejä mesimarjakantoja*. Yhtään sellaista *ei löytynyt*. Mesimarja osoittautui siis Suomessa varsin vallitsevasti itsesteriiliksi lajiksi. Sikäli kuin itsefertilejä kantoja lainkaan esiintyy, ne ainakin olisivat harvinaisia poikkeuksia.

Tutkimuksessa osoitettiin, että inkompatibiliteetti luokkien *lukumäärällä* on yleensä ratkaiseva merkitys mesimarjan hyväkuntoisien luonnonesiintymien marjonnalle. Lähes aina tällaisen mesimarjapopulaation *marjomattomuus* johtui siitä, että siinä esiintyi vain *yhtä* inkompatibiliteetti luokkaa olevia versoja. Seitsemässä tapauksessa kahdeksasta marjomattomasta esiintymästä löytyi vain yhtä inkompatibiliteetti luokkaa. Vain yksi marjomaton populaatio oli useampiluokkainen; sen marjomattomuuden lienee aiheuttanut jokin luonnonkasvupaikan ympäristötekijä (esim. pölyttäjien puute), sillä koekentällä tämäkin populaatio marjoi normaalisti. Yksi populaatio osoittautui triploidiseksi eikä kyennyt tuottamaan marjoja minkäänlaisissa risteytyksissä.

*Runsaasti marjovissa* populaatioissa esiintyi yleensä kolmesta seitsemään inkompatibiliteetti luokkaa. Vieraiden kukkien osuus nousikin näissä yleensä varsin korkeaksi (välillä .42—.86), joten hedelmöittämisselkelpoisen siitepölyn saatavuus oli hyvä. *Heikosti* marjovissa esiintymissä oli kahdesta kuuteen inkompatibiliteetti luokkaa. Eräissä esiintymissä vieraiden kukkien osuus jäi pieneksi, eräissä taas populaation alueellinen rakenne ei suosinut luokkien välistä pölytystä. Usein myös jokin ympäristötekijä (esim. muiden lajien kilpailu) oli jo epäedullinen — heikosti marjovia populaatioita ei yleensä voitukaan pitää täysin hyväkuntoisina.

Ukottavimman tulkinnan mukaan yksiluokkaiset esiintymät ovat kokonaan *yhtä yksilöä*, joka yhdestä sieme-

nestä alunperin lähteneenä on juurakon avulla levittänyt laajallekin alueelle. Tosin vielä kaksi- tai kolmisiemeninenkin alkuperä olisivat melko uskottavia selityksiä. Suurin tutkimuksessa löytynyt mesimarjaklono on siis halkaisijaltaan noin 80 metriä, mikä vastanee ainakin noin 160 vuoden ikää. Tällainen pitkäikäisyys antaisi tukea kasvullisen ryöpsähdyksen käsitykselle. Sen mukaan mesimarjakasvustoa säilyisi — joskin huonokuntoisena ja huomaamattomana — varsin pitkään entisillä esiintymispaikoillaan. Kun olot jälleen muuttuvat suotuisammiksi, mesimarja pystyisi tällöin tehokkaan juurakostonsa avulla parissa-kolmessa vuodessa taas täyttämään alueen tiheällä kavustolla. Mikäli maassa on hautautuneena mesimarjan siemeniä entisiltä kukoistuskasuilta — ehkä vuosikymmenienkin takaa — täytynee kuitenkin myös *sie-menten itämisellä* olla toisinaan osuutensa ryöpsähdyksissä.

Tämän tutkimuksen läpiviemiseksi jouduttiin kehittämään joitakin *menetelmällisiä uutuuksia*. Risteytystyön vähentämiseksi laadittiin tietokoneistettu risteytysten ohjauksjärjestelmä RISTO. Työ järjestettiin tapahtuvaksi kierroksittain. Kullakin kierroksella järjestelmä analysoi aina kaikki siihen asti luetut tulokset ja antoi suosituksia uusiksi risteytyksiksi, joiden avulla todennäköisesti saataisiin eniten lisätietoa.

Populaation yksilöiden ryhmittäminen *inkompatibiliteettiluokiksi* arvioitiin kertyneiden risteytystulosten pohjalta soveltaen suurimman uskottavuuden periaatetta. Uskovimman luokkajaon löytämiseksi  $8.47 \cdot 10^{23}$  mahdollisesta, laadittiin laajennettu yksittäissiirtomenetelmä, jossa ryhmittämisestä toiseen edettiin kerrallaan vain yhden yksilön ryhmittämisellä tai tarvittaessa kaksi ryhmää yhdistämällä.

Ryhmittelyn perustana oli risteytyksen tulos: siemenlukumäärä (joka käsitettiin *satunnaisuuttujaksi*). Vastaavuusluokitus oli mahdollista, koska inkompatibiliteettiluokan sisäisissä risteytyksissä siemenlukumäärän todennäköisyysjakautuma oli erilainen kuin luokkien välisissä. Siemenlukumäärien todennäköisyysjakautumat estimoitin *ei-parametrisesti*, käyttäen sopivasti muunnettua Parzenin kernel-menetelmää. Todennäköisyysjakautumia ja inkompatibiliteettiluokkajakoa *tarkennettiin* lopuksi yhdessä *iteroimalla*. Iteraatio suppeni aina nopeasti: jo 4.—7. kierroksella.

Risteytysuusitusten laatimiseksi kehitettiin menetelmä, joka nojasi suurin labiilisuus — suurin stabiilisuus -periaatteeksi nimeämäni strategiaan. Täydellistä todistusta sille, että tällainen strategia välttämättä johtaa mahdollisimman tehokkaisiin risteytyksiin, ei voitu esittää — ei myöskään luotettavaa arviota saavutettavan säästön suuruudesta. Menetelmän teoreettinen perustelu jäi siis osaksi heuristiselle tasolle. Käytännön tulosten perusteella suosituksia kuitenkin osoittautuivat ilmeisen *tehokkaiksi*. Risteytysten ohjaussysteemin (RISTO) avulla selvittiin noin 10 prosentilla siitä risteytysten määrästä, joka olisi vaa-

dittu tavanomaista (puolialleliristeytys)menetelmää sovellettaessa yksinkertaisempaan (deterministisen muuttujan) tapaukseen. Väitöskirjatyössä tarvittiin siten kaikkiaan vain noin 1200 (netto)risteytystä.

Yksiluokkainen, marjomaton luonnonesiintymä alkaneen yleensä *marjoja*, jos vain järjestetään vieraan, hedelmöittämissä siitepölyn saanti. Eri puolille esiintymää voidaan esim. istuttaa jotain *toista*, muualta löydettyä tai vaikkapa ostettua mesimarjakantaa. Myös jalomaarain käy periaatteessa pölyttäjäksi. Istukkaille on varattava myös kasvutilaa, jotta vieraiden kukkien osuus voi nousta riittäväksi.

Mesimarjaviljelmälläkin tulisi marjonnan varmistamiseksi kasvattaa mieluiten vähintään kolme eri mesimarjakantaa. Mitä useampaa kantaa viljelmällä esiintyy, sitä suurempi on vieraiden, hedelmöittämissä kukkien osuus. Lisäksi pölyn tehokas kulkeutuminen kannasta toiseen on syytä turvata istuttamalla kantoja hyvin *sekaisin* kuhunkin riviin — ei siis vuororiveiksi kuten ennen suositeltiin. Tällaisella viljelmällä satotaso ei heti romahtaisi, jos yhdessä mesimarjakannassa esiintyy vaikeuksia (esim. kasvu on heikkoa tai kukinta tärvytys tautien, loisten tai muiden satunnaisten tekijöiden vuoksi).

Lähdettäessä jalostamaan luonnonkasvia, on ensimmäisenä tehtävänä *laajan geneettisen kokoelman* perustaminen. Tämä on tarpeen jo siksi, että jalostustavoitteiden joukko on laaja ja epämääräinen. Viljelymenetelmät ja sadon käyttötavatkin ovat nimittäin vielä käymistilassa (esim. vapaassa rivissä, muovikatteen rajaamana vai riippusäkeissä viljely mesimarjalla), ja uudet menetelmät saattavat edellyttää kasvilta aivan erilaisia ominaisuuksia kuin aiemmat. Lisäksi valinta suoraan luonnosta on jalostuksen kannalta suhteellisen tehontonta, koska a) vaihtelevassa luonnonympäristössä ominaisuuksien periytyvyys on pieni ja b) viljely-ympäristö on toinen kuin valintaympäristö. Jalostuksellista *valintaa* päästään toden teolla harjoittamaan vasta sellaisessa laajassa kokoelmassa, joka on perustettu riittävän tasalaatuisiin ja todellista viljelymenetelmää oleellisesti vastaaviin oloihin.

Tämän tutkimuksen perusteella yhdessä mesimarjapopulaatiossa yleensä kasvaneen enintään muutamaa genotyyppiä. Jalostuskokoelmaa kerättäessä lienee siis voimavarat tehokkainta suunnata siten, että näytteitä saataisiin mahdollisimman *monesta* eri populaatiosta — muutama näyte kustakin taas riittänee.

Kaus vaeltava, harvahko ja helposti limittyvä juurakon *kasvutapa* soveltui hyvin mesimarjan luonnonpopulaatioihin. Viljelyoloja ajatellen sitä kannattanee pyrkiä jalostuksella *lyhentämään* — ainakin jos mesimarjaa ryhdytään viljelemään muovikate- tai riippusäkkimenetelmällä. Kun sekä ravinteiden, valon että pölyttäjäkantojen saanti on ihmisen toimin turvattu, juurakoitten kasvataminen kaus olisi kasvin resurssien tuhlausta.

Viljelytekniseltä kannalta myös *itsefertiilisyys* olisi tär-

keä jalostustavoite — silloin mesimarjaviljelmällä tul-tai-siin toimeen yhdellä lajikkeella. Polyploidia- tai mutaa-tiojalostuksella tavoitteeseen saatettaisiin päästä. Itsefer-tiilin mesimarjakannan etsimistä luonnostakaan ei kan-nattane vielä lopettaa. Onhan viljellyllä vadelmallakin it-sefertiiisyyttä yleistä, viljelyn kuluessa tapahtuneen tiedo-tamattoman valinnan tuloksena. Itsesiitoksen seurauksena esiintyisi tällaisella luonnostaan tiukasti ristisiitteisellä kas-vilajilla voimakas *sisäsiitostaantuma*. Jos itsestään aiot-taisiin jossain muodossa ryhtyä käyttämään hyväksi me-simarjan jalostusohjelmissä, kannattaisi siis uuden jalos-tustekniikan keinoin (haploidiviljely ja kromosomiston kaksinkertaistaminen) heti alussa *massamitassa* tuottaa täydellisesti *homotsygoottisia* mesimarjakantoja. Näin saataisiin jalostuspopulaatioista karsituiksi pois haitalli-set (letaalit ja subvitaalit) alleelit: homotsygoottisuuden ansiosta sellaisia sisältävät yksilöt kuolisivat tai ne olisi huonokasvuisina helppo karsia jo varhaisessa vaiheessa. Homotsygootteja täytyisi valmistaa paljon, sillä elinvoi-maisiksi niistä saattaisi alkuvaiheessa osoittautua ehkä vain aniharva kanta. Haitallisten alleelien karsiuduttua — kuten "vanhoilla" itsestään lajeilla on evoluution kuluessa tapahtunut — sisäsiitostaantuma ei jatkossa enää olisi kiusana.

Eräät tutkimuksen kuluessa esiin nousseet kysymykset ansaitsisivat vielä jatkotutkimuksia. Mesimarjan *kloonien kokoa* luonnonesiintymässä tulisi tutkia esim. DNA-sor-menjälkimenetelmällä (JEFFREYS ym. 1985a, b). Mesimar-jan siementen mahdollinen säilyminen *itämiskelpoisina* maassa vuosien ja vuosikymmenten ajan kaipaisi kokeel-lista selvitystä. Esiintymispaikan maaperän siemenfloo-ran tutkiminen voisi valaista erityisesti yksiluokkaisten me-simarjapopulaatioiden *taustaa*: onko paikalla aiemmin esiintynyt marjova populaatio? Mesimarjan inkompati-biliteettialleelien *kokonaismäärä* Suomessa tulisi arvioi-da. Samoin kannattaisi selvittää eri inkompatibiliteetti-alleelien ja -genotyyppien *alueellista jakautumaa* maas-samme; esiintyykö kenties vyöhyke-eroja? Risteytystyön ohjauksella saavutettavan *säästön* arvioimiseksi tarvittai-siin teoreettista jatkotutkimusta. Ohjaussysteemin ylei-sen *luotettavuusasteen* arvioimiseksi tulisi tehdä vielä lisäkokeita uusilla aineistoilla. Itse asiassa RISTO on periaatteessa *yleinen* menetelmä: sen pitäisi soveltua eri-laisten vastaavuusluokitteluun tähtäävien kokeiden op-timaaliseen ohjaamiseen. Sitä voitaisiin siis kokeilla myös *muilla tieteenaloilla*.

# 11. Appendices

## App. A. Arctic bramble experimental field at Viikki in 1977—80

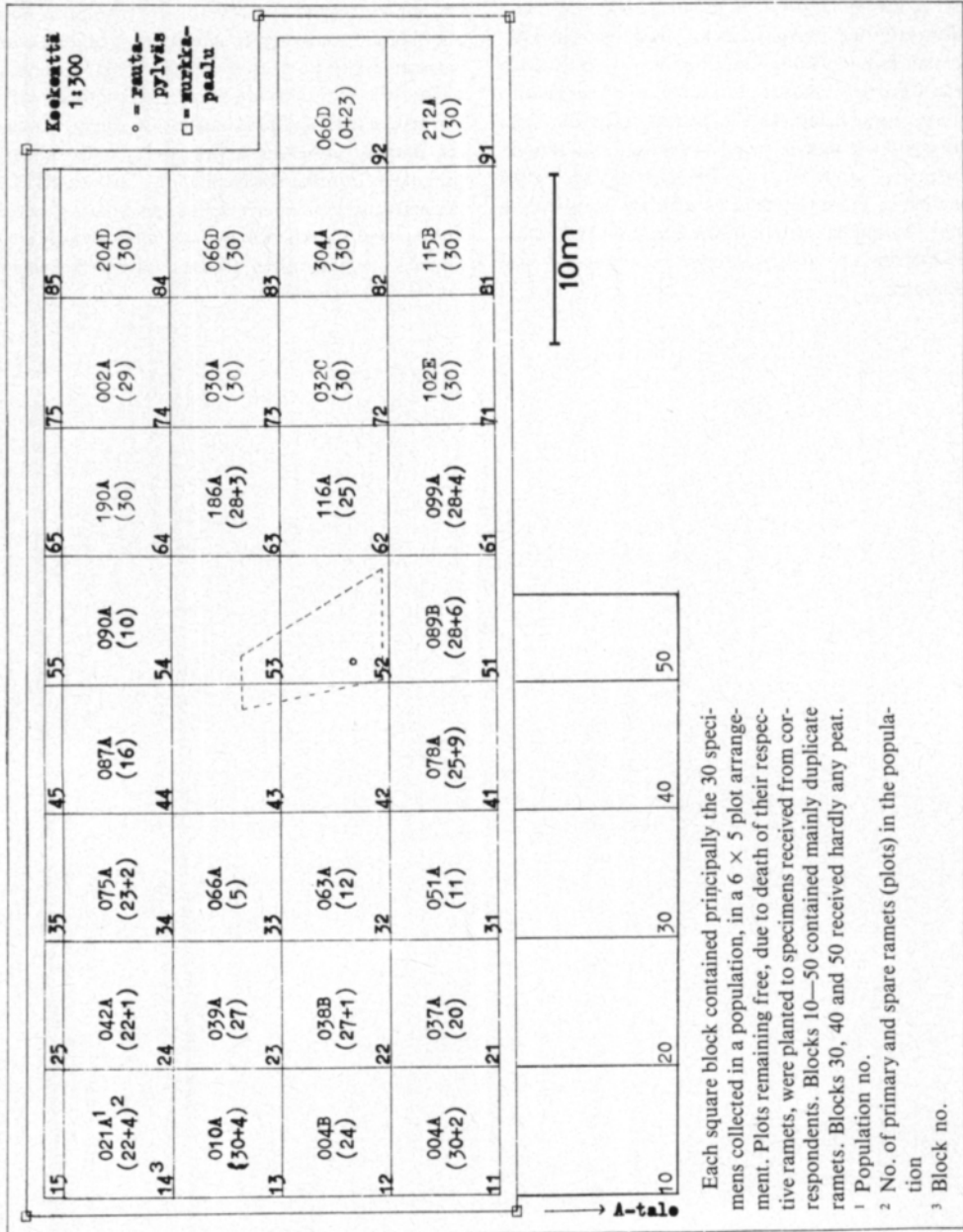
KASVINJALOSTUS-  
TIETEEN LAITOS

### SELOSTE

Systeemi		Mesimarjatuotus	
Laatinut	02.05.79 15.05.76	Osa	Alakohta
JTa	27.05.77		Sivu

Kuvauskohde

Istutus - yleiskartta



Each square block contained principally the 30 specimens collected in a population, in a 6 x 5 plot arrangement. Plots remaining free, due to death of their respective ramets, were planted to specimens received from correlative respondents. Blocks 10—50 contained mainly duplicate ramets. Blocks 30, 40 and 50 received hardly any peat.

- <sup>1</sup> Population no.
- <sup>2</sup> No. of primary and spare ramets (plots) in the population
- <sup>3</sup> Block no.



# App. B. Crossing form

Ripono 10 Ripom  
Rinl6 kk v Ril6  
102EJ190679058U  
66 67 68 69 70 71 72 73 74 75 76 77 78 79 80  
N N N A A A A N N N N N N N N N A

## RISTEYTYKSET

Rubus arcticus L.

Rino <sup>1</sup>	Eykno <sup>2</sup>	Hykno <sup>3</sup>	Ripno <sup>5</sup>	Pikno <sup>6</sup>	Epono <sup>7</sup>	Hpono <sup>8</sup>	Rihuk <sup>4</sup>	Rihuk <sup>4</sup>	Rihuk <sup>4</sup>
N N A N N A	N N A N N A	N N A N N A	N N A N N A	N N A N N A	N N A N N A	N N A N N A	N N A N N A	N N A N N A	N N A N N A
1. 2. 3. 4. 5. 6. 7	8. 9. 10. 11. 12. 13. 14.	15. 16. 17. 18. 19. 20. 21. 22. 23. 24.	25. 26. 27. 28. 29. 30. 31. 32.	33. 34. 35. 36. 37. 38.					
005205	20	24	23	0052099A					
005324	20	23							
005414	20	24	23						
005504	20	23		0054099A					
005620	25	24	23						
005704	25	23		0056099A					
005820	13	24	23			55	61		
005912	13	23		0058186A		55			
006021	20	23	23			15			
006113	20	23		0060186A					
006224	03	24	23						
006327	03	23		0062099A					
006403	09	24	23			55			
006530	09	23		0064099A		55			

Risteytyshuomautusten (Rihuk1, ..., Rihuk4) koodit 9 (JtA 20.05.79)

- Hedekukka Huo! Märkä siitepölyä ei saa käyttää!
- 05 Pöly enalluokk, ja sitä herv, runs.
- 51 Pöly vähän muuta kunnoll. (vaata suonn.)
- 52 Hedekk. härmsäinen
- 54 Hedekk. lyhytteräl., kalspöb.
- 55 Hedekk. emittön
- 02 Pöly liian nuorta (oli irr, eksikkaast.)
- 50 Pöly vanhaa: valk., varisevaa; ponaissa rusketusta
- 07 " oik. laidassa
- 08 " " takalaidassa
- 53 Hedekk. näivetytä tai paljon ripsiäisiä
- 57 " " vas. laidassa
- 04 Petriemajassa ripsiäisiä
- 09, 59 var. onatek, huomautuksille

- Eiikkukka
- 10 Emejä vähän mutta kunnollisia
- 11 Nuppu liian nuori: emitt pientä,
- 60 Nuppu liian nuori: emitt hyvin pientä, pöly ei tartu
- 15 Emitt pientä vaikka kk, jo suk.vaiheessa
- 12 Eiikk. liian vanha: terälehtiä irtoillee
- 61 Eiikk. liian vanha: leikkatessa luotit kastuvat neodestä
- 64 Eiikk. lyhytterälehtinen, kälkioimäinen
- 65 Emitt epäsuodostuneet: vinot tas
- 62 Eiikkussa ripsiäisiä
- 63 Eiikk. härmsäinen
- 66 Vertailot tai luotit ruskett.
- 14 Sikäimet ruakaita
- 13 Nuppu reisiäinen nuorena
- 16 Eiikk. ruudun etulaidassa
- 17 " " oik. laidassa
- 18 " " takalaidassa
- 67 " " vas. laidassa
- 19, 69 var. onatek, huomautuksille

- Risteytystekniikka
- 23 Eiikkukana eristetty nuppu
- 24 Eiikkukana nuori, suennut kukka (eristetty nupuna)
- 25 Käytettiin adosuojuusta
- 26 Pöly irrotettiin laeulle
- 22 Pöly irrotettiin eksikkaatorissa
- 72 Pöly kuvunnet (kruun auringosaa tai eksikkaatorissa, calhot koooraisia)
- 70 Eiikk. jo liian avoin eristämön nuppu
- 20 Eiikk. onaskul. katk. heheet
- 71 Meisi poisit, eiikkukaata vesi-huunt., kulv. puhaltaaalla
- 21 Eiikk. erist. oikk. nuorena
- 73 Sade alkoi alle 3t päältä, ei käyt. adosuojuusta
- 78 Risteyttäjä harjoitteluvaih.
- 29, 79 var. onatek, huomautuksille

Onatekkoiset huomautukset (Pätevät vain yden louskkaan sjan! Selitye kirjoitettava!)

- 09
- 59
- 19
- 69
- 29
- 79
- 1 Cross no.
- 2 Seed parent no.
- 3 Pollen parent no.
- 4 Comment
- 5 Crossing date
- 6 Pollen flower no.
- 7 Population of seed parent
- 8 Population of pollen parent
- 9 Comment codes
- 10 Investigator

App. C. Recording prescriptions, and cross results recorded

RUBUS ARCTICUS L.												RESULTS (TO BE) RECORDED					LISTED								
CO-ORD.	SEED	CROSS	NO.	NO.	DAY	NO.	M.	PDLL.	NO.	NO.	CD.	CD.	CD.	CD.	NO.	NO.	CD.	CD.	CD.	CD.	CD.				
23.06	34-5.2	>075A-30	1	0053	01	26	.06	---	075A-21	01	---	---	---	---	---	---	---	---	---	---	---	---	---	---	
23.06	34-6.1	>130C-00V	44	075A-0056	44	26	.06	---	075A-04	44	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
23.06	34-6.1	>130C-00V	45	075A-0056	45	26	.06	---	075A-21	45	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
23.06	34-6.3	>135A-00	08	000K-0056	08	26	.06	---	066A-13	08	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
21.06	35-1.1	>136C-00	26	042A-0076	26	26	.06	---	042A-02	26	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
21.06	35-1.1	>136C-00	36	042A-0078	36	26	.06	---	042A-25	36	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
22.06	35-3.1	>137B-00	39	021A-0059	39	26	.06	---	021A-01V	39	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
22.06	35-3.1	>137B-00	35	075A-0044	35	26	.06	---	075A-10	35	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
22.06	35-3.1	>137B-00	28	075A-0046	28	26	.06	---	075A-13	28	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
23.06	35-4.1	>142A-00	28	004A-0243	28	26	.06	---	004A-18	28	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
23.06	35-4.1	>142A-00	42	075A-0054	42	26	.06	---	075A-21	42	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
22.06	40-1.2	>212A-12V	00	0032	00	26	.06	---	212A-16V	00	48	---	---	---	---	---	---	---	---	---	---	---	---	---	---
23.06	40-1.4	>212A-14V	00	0036	00	26	.06	---	212A-16V	00	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
21.06	40-1.6	>212A-16V	00	0026	00	26	.06	---	212A-22V	00	47	---	---	---	---	---	---	---	---	---	---	---	---	---	---
22.06	40-1.6	>212A-16V	00	0030	00	26	.06	---	212A-11V	00	48	---	---	---	---	---	---	---	---	---	---	---	---	---	---
23.06	40-1.6	>212A-16V	32	0038	32	26	.06	---	212A-17V	32	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

1 Recommended date of recording  
2 Field co-ordinate of the seed parent  
3 Seed parent no.  
4 Resulting no. of seeds (written twice to minimise risk of poor deciphering of handwritten observations)  
5 Month  
6 Pollen parent in the cross  
7 Comment codè in recording  
8 Recorder's initials (blank denotes JTa)  
9 Initials of performer of the crosses  
10 Page no. of the crossing form  
11 Date of crossing

App. D. Estimated equivalence classes of populations

POPULAAATIO 032C

OSOITETUT EKVIVALENSSILUOKAT<sup>1</sup>

RIST•TARVE<sup>2</sup> EKV•L<sup>3</sup> YKSILÖT VAHVIMMASTA LXHTIEN<sup>4</sup>

-•290U+01 ( 6)	24	11	20	25	28	29V	23V	12	01	22	21	19	18V	14	13	10V	09	05	04
-•150U+01 ( 3)	08	03V	26																
-•155U+01 ( 4)	16V	15	07																
0• ( 7)	27V																		

ERILLISYYDET<sup>5</sup>

	24	08	16V	27V
( 6)	( 3)	( 4)	( 7)	( 7)

NÄYHÄISET EKVIV•LUOKAT<sup>6</sup> (=PUUTTEELLISESTI TUTKITUT YKSILÖT)

0• ( 8)	30	1•4	0•6	1•1	0•0
0• ( 2)	06	1•3	0•0	0•0	0•0
0• ( 5)	17V	1•4	1•3	0•0	0•0
0• ( 1)	02V	0•0	1•4	0•0	0•0

POPULAAATION 032C YKSILÖIREN SAMUUSARVOT ERI EKVIVALENSSILUOKKIIN<sup>7</sup>

FKVIV•LUOKKA 9 = 'IKRALLISEKSI' YKSILÖT OVAT EKVIV•LUOKITTAIN, VAHVIMMAN EKVIV•LUOKAN VAHVIMMASTA YKSILÖSTÄ LXHTIEN.

YKSILÖ (EKVIVSJJ)	(EKVIVSJJ)	(EKVIVSJJ)	(EKVIVSJJ)	(EKVIVSJJ)	(EKVIVSJJ)	(EKVIVSJJ)	(EKVIVSJJ)	(EKVIVSJJ)	(EKVIVSJJ)	(EKVIVSJJ)	(EKVIVSJJ)	(EKVIVSJJ)	(EKVIVSJJ)	(EKVIVSJJ)	(EKVIVSJJ)	(EKVIVSJJ)	(EKVIVSJJ)	(EKVIVSJJ)	(EKVIVSJJ)	(EKVIVSJJ)
24 0 ( 6)	- { 1680+02	- { 1580+02	- { 1680+02	- { 1680+02	- { 1680+02	- { 1680+02	- { 1680+02	- { 1680+02	- { 1680+02	- { 1680+02	- { 1680+02	- { 1680+02	- { 1680+02	- { 1680+02	- { 1680+02	- { 1680+02	- { 1680+02	- { 1680+02	- { 1680+02	- { 1680+02
11 0 ( 6)	- { 7550+01	- { 7550+01	- { 7550+01	- { 7550+01	- { 7550+01	- { 7550+01	- { 7550+01	- { 7550+01	- { 7550+01	- { 7550+01	- { 7550+01	- { 7550+01	- { 7550+01	- { 7550+01	- { 7550+01	- { 7550+01	- { 7550+01	- { 7550+01	- { 7550+01	- { 7550+01
20 0 ( 6)	- { 6090+01	- { 6090+01	- { 6090+01	- { 6090+01	- { 6090+01	- { 6090+01	- { 6090+01	- { 6090+01	- { 6090+01	- { 6090+01	- { 6090+01	- { 6090+01	- { 6090+01	- { 6090+01	- { 6090+01	- { 6090+01	- { 6090+01	- { 6090+01	- { 6090+01	- { 6090+01
25 0 ( 6)	- { 3780+01	- { 3780+01	- { 3780+01	- { 3780+01	- { 3780+01	- { 3780+01	- { 3780+01	- { 3780+01	- { 3780+01	- { 3780+01	- { 3780+01	- { 3780+01	- { 3780+01	- { 3780+01	- { 3780+01	- { 3780+01	- { 3780+01	- { 3780+01	- { 3780+01	- { 3780+01
28 0 ( 6)	- { 3470+01	- { 3470+01	- { 3470+01	- { 3470+01	- { 3470+01	- { 3470+01	- { 3470+01	- { 3470+01	- { 3470+01	- { 3470+01	- { 3470+01	- { 3470+01	- { 3470+01	- { 3470+01	- { 3470+01	- { 3470+01	- { 3470+01	- { 3470+01	- { 3470+01	- { 3470+01
29V 0 ( 6)	- { 1460+01	- { 1460+01	- { 1460+01	- { 1460+01	- { 1460+01	- { 1460+01	- { 1460+01	- { 1460+01	- { 1460+01	- { 1460+01	- { 1460+01	- { 1460+01	- { 1460+01	- { 1460+01	- { 1460+01	- { 1460+01	- { 1460+01	- { 1460+01	- { 1460+01	- { 1460+01



23V	0 ( 6 )	- ( 1 )	- ( 2 )	- ( 3 )	- ( 4 )	- ( 5 )	- ( 7 )	- ( 8 )	- ( 9 )
12	0 ( 6 )	- ( 1 )	- ( 2 )	- ( 3 )	- ( 4 )	- ( 5 )	- ( 7 )	- ( 8 )	- ( 9 )
01	0 ( 6 )	- ( 1 )	- ( 1 )	- ( 2 )	- ( 5 )	- ( 6 )	- ( 8 )	- ( 9 )	- ( 3 )
22	0 ( 6 )	- ( 1 )	- ( 1 )	- ( 3 )	- ( 5 )	- ( 6 )	- ( 8 )	- ( 9 )	- ( 4 )
21	0 ( 6 )	- ( 1 )	- ( 1 )	- ( 3 )	- ( 4 )	- ( 5 )	- ( 7 )	- ( 8 )	- ( 9 )
19	0 ( 6 )	- ( 1 )	- ( 1 )	- ( 3 )	- ( 4 )	- ( 5 )	- ( 7 )	- ( 8 )	- ( 9 )
18V	0 ( 6 )	- ( 1 )	- ( 1 )	- ( 3 )	- ( 4 )	- ( 5 )	- ( 7 )	- ( 8 )	- ( 9 )
14	0 ( 6 )	- ( 1 )	- ( 1 )	- ( 3 )	- ( 4 )	- ( 5 )	- ( 7 )	- ( 8 )	- ( 9 )
13	0 ( 6 )	- ( 1 )	- ( 1 )	- ( 3 )	- ( 4 )	- ( 5 )	- ( 7 )	- ( 8 )	- ( 9 )
10V	0 ( 6 )	- ( 1 )	- ( 1 )	- ( 3 )	- ( 4 )	- ( 5 )	- ( 7 )	- ( 8 )	- ( 9 )
09	0 ( 6 )	- ( 1 )	- ( 1 )	- ( 3 )	- ( 4 )	- ( 5 )	- ( 7 )	- ( 8 )	- ( 9 )
05	0 ( 6 )	- ( 1 )	- ( 1 )	- ( 3 )	- ( 4 )	- ( 5 )	- ( 7 )	- ( 8 )	- ( 9 )
04	0 ( 6 )	- ( 1 )	- ( 1 )	- ( 3 )	- ( 4 )	- ( 5 )	- ( 7 )	- ( 8 )	- ( 9 )
16V	0 ( 4 )	- ( 2 )	- ( 2 )	- ( 3 )	- ( 9 )	- ( 6 )	- ( 3 )	- ( 8 )	- ( 7 )
15	0 ( 4 )	- ( 1 )	- ( 2 )	- ( 5 )	- ( 8 )	- ( 9 )	- ( 1 )	- ( 3 )	- ( 6 )
07	0 ( 4 )	- ( 1 )	- ( 1 )	- ( 5 )	- ( 6 )	- ( 8 )	- ( 9 )	- ( 3 )	- ( 6 )
08	0 ( 3 )	- ( 2 )	- ( 9 )	- ( 2 )	- ( 3 )	- ( 7 )	- ( 5 )	- ( 6 )	- ( 6 )
03V	0 ( 3 )	- ( 1 )	- ( 4 )	- ( 5 )	- ( 7 )	- ( 8 )	- ( 9 )	- ( 6 )	- ( 6 )
26	0 ( 3 )	- ( 2 )	- ( 1 )	- ( 2 )	- ( 5 )	- ( 6 )	- ( 7 )	- ( 8 )	- ( 9 )
30	0 ( 8 )	0 ( 7 )	0 ( 2 )	0 ( 5 )	0 ( 1 )	- ( 3 )	- ( 4 )	- ( 6 )	(
27V	0 ( 7 )	0 ( 8 )	0 ( 2 )	0 ( 5 )	0 ( 1 )	- ( 3 )	- ( 4 )	- ( 6 )	(
17V	0 ( 5 )	0 ( 4 )	0 ( 2 )	0 ( 1 )	0 ( 7 )	0 ( 8 )	- ( 3 )	- ( 6 )	(
06	0 ( 2 )	0 ( 4 )	0 ( 1 )	0 ( 5 )	0 ( 7 )	0 ( 8 )	- ( 3 )	- ( 6 )	(

02V 0 ( 1 ) U ( 6 ) ^ ( 4 ) 0 ( 5 ) 0 ( 2 ) 0 ( 7 ) 0 ( 8 ) - 1360+01  
 8 RYHALT=9 2142U+01 AKLPKH=2 NCSAMJ=2 NCERTIJ=2  
 ANKLPJ=57 ANHYLJ=11E POTILJ=175 RYSILJ=10 RYHHVHJ=11 YJXSLJ=2  
 -ANADUT:N RIVIKIRJUITJULOSTUS -  
 RIPONU=032C, NJ=30, NLJ= 8  
 NJE= 1, 1, 3, 3, 1, 19, 1, 1,  
 ANAYKJ= 01, 02V, 03V, 04, 05, 06, 07, 08, 09, 10V, 11, 12, 13, 14, 15, 16V, 17V, 18V, 19, 20, 21, 22, 23V, 24, 25, 26, 27V, 28, 29V, 30,  
 RITRVJ= -2, 1423  
 KERILI= 0, 19392, 0, 19237, 1, 8840, 1, 8391, 0, 37673, 2, 6129, 0, 65316, 0, 43508,  
 RIV:\*\* 0, 1, 4969, 1, 5454, 0, 2, 9022, 0, 0,  
 RITRJI= 0, 1, 4969, 1, 5454, 0, 2, 9022, 0, 0,  
 RYHALT=1, AKLPKH=2, NCSAMJ=2, NCERTIJ=2, ANKLPJ=57, ANHYLJ=118, RYHHVHJ=1,  
 RYSILJ=22, YJXSLJ=2, POTILJ=175

-ANADUT:N RIVIKIRJUITJULOSTUS (RIPONU=032C) PÄÄTTY -

- 1 Estimated equivalence classes
- 2 Average lability (Lab<sup>(n)</sup> (e<sub>n</sub>)) of ramets in an equivalence class (e<sub>n</sub>) (see expression (80) in 3.3.5.7)
- 3 Equivalence class no.
- 4 Ramets of an equivalence class, in increasing order of lability
- 5 Pairwise dissimilarities (Diss<sup>(n)</sup> (e<sub>n</sub>, e<sub>n</sub>)) between estimated equivalence classes (e<sub>n</sub> and e<sub>n</sub>) (see expression (82) in 3.3.5.7)

- 6 Unclassified or partially classified ramets
- 7 Standardized similarities (S<sub>std</sub><sup>(n)</sup> (r, e<sub>n</sub>)) of the ramets (r) to equivalence classes (e<sub>n</sub>) (see expression (76) in 3.3.5.5)
- 8 Current average lability (Lab<sup>(n)</sup>) of the ramets in the population (see expression (81) in 3.3.5.7)
- 9 Net no. of crosses (see Tab. 5)
- 10 No. of single moves (see Tab. 5)
- 11 No. of amalgamations (see Tab. 5)

App. E. Crossing recommendations

DATE 25.06.80 P. 1

-1.09

CROSSES PROPOSED  
NEED OF CROSSING

PUBUS ARCTICUS L.  
POPULATION '021A'

MADE	1.PARENT	2.PARENTS	CROSSES PROPOSED	NEED OF CROSSING
01V	14-5.00	26 14-5.00		
25	14-4.54	05 14-2.07		
26	14-5.1	08 14-2.02		
05	14-1.5	14.8 14-2.05		
06	14-2.1	09 14-2.03		
08	14-2.2	26 14-5.1		
09	14-2.3	12 14-2.04		
12	14-2.4	09 14-2.03		
14	14-2.5	18 14-3.03		
17	14-3.2	09 14-2.03		
18	14-3.3	14 14-2.05		

MADE	1.PARENT	2.PARENTS	CROSSES PROPOSED	NEED OF CROSSING
19	14-1.16	15 10-3.5 14-3.1	22 14-1.4 04	24 14-5.2 28 14-1.46 03 14-1.94 22
24	14-1.16	15 10-3.5 14-3.1	22 14-1.4 04	21 14-1.16 19 14-3.16 19 14-1.16 21 14-1.16 21 14-1.16
21	14-1.16	15 10-3.5 14-3.1	22 14-1.4 04	24 14-5.2 28 14-1.46 03 14-1.94 22
28	14-1.16	15 10-3.5 14-3.1	22 14-1.4 04	24 14-5.2 28 14-1.46 03 14-1.94 22
03	14-1.46	15 10-3.5 14-3.1	22 14-1.4 04	24 14-5.2 28 14-1.46 03 14-1.94 22
22	14-1.42	15 10-3.5 14-3.1	03 14-1.46 24	21 14-1.16 19 14-3.16 19 14-1.16 21 14-1.16
04	14-1.14	15 10-3.5 14-3.1	22 14-1.42 03	24 14-5.2 28 14-1.46 03 14-1.94 22

DATE 25.06.80 P. 2

-1.09

CROSSES PROPOSED  
NEED OF CROSSING

PUBUS ARCTICUS L.  
POPULATION '021A'

MADE	1.PARENT	2.PARENTS	CROSSES PROPOSED	NEED OF CROSSING
15	10-3.5 14-3.1	04 14-1.4	22 14-1.42 03	24 14-5.2 28 14-1.46 03 14-1.94 22
19	14-1.16	15 10-3.5 14-3.1	24 14-1.16 19	21 14-1.16 21 14-1.16 21 14-1.16

1 An asterisk (\*) denotes that at least one of the recommended crosses has already been made (but the results have not been recorded yet)  
 2 No. of proposed first parent (in descending order of recommendation)  
 3 Liability (Lab<sup>m</sup> (r)) of the first parent between its 'own' class and the class 'most strongly rivalling' it (see expressions (76) in 3.3.5.5 and (79) in 3.3.5.7)  
 4 Field co-ordinate of the first parent  
 5 Proposed second parent (recommended order is from left to right)

6 Liability (Lab<sup>m</sup> (r)) of the second parent between its 'own' class and the class 'most strongly rivalling' it (see expressions (76) in 3.3.5.5 and (79) in 3.3.5.7)  
 7 Field co-ordinate of the second parent  
 8 An underlying row of asterisks (\*\*\*) denotes that a pending cross (see 1 above) has been made with the second parent in question  
 9 Current average liability (Lab<sup>m</sup>) of the ramets in the population (see expression (81) in 3.3.5.7)

NO. OF CROSSING RECOMMENDATIONS

## App. F. Details of the crossing technique

To avoid contamination, hands were washed and equipment sterilized with 96 % ethanol between two crosses (70 % would have been more efficient, however). Care was also taken, not to touch the pistils by hand or forceps but only with the pollen sacks of the relevant pollen flower. Though no longer attractive to insects, the pollinated flowers were immediately enclosed in another bag — made out of woven cotton fabric — and labeled. Both types of bags were tightened by wrapping cotton wadding round the flower stalk, and closed by winding coloured plastic tape around the neck. In spite of care taken, thrips were occasionally found on isolated flowers. Almost always such flowers were abandoned. Any reason for uncertainty was always carefully recorded, as well as any other exceptional details. Hence, if a flower with uncertain purity was sometimes obliged to be used in crosses — due to a severe shortage of flowers in the clone — their results could be left out of analyses as soon as more relevant crossing data became available.

About ten to fourteen days after the cross was made, the results were recorded. In the pollinated flower, the number of ovaries beginning to enlarge was counted. In a destructive examination — cutting the flower stalk and breaking the infantile fruit under a binocular microscope in the laboratory — one could have also counted the number of non-enlarging ovaries. However in order to save the seed of these crosses, a non-destructive method was chosen. The bag was opened and the enlarging ovaries were counted *in situ*, marking each with a tiny spot with an indelible marker. The bag was then replaced, to be harvested some weeks later with a ripe berry in it.

The computer system (see Fig. 8) provided in advance a list of the crosses to be inspected (App. C). This list was a sorted one, i.e. arranged according to the map position of the seed parent in the experimental field, and minimized walking back and forth. The list itself served as a results form in the field. After being filled up, the same form was to be used to transfer the cross results data into the memory of the computer. The computer generated form simplified note keeping in the field, and minimized mistakes due to human errors in writing, or to inconsistent handwriting.

## App. G. Equivalence relation and equivalence classes

A relation  $R$  is defined to be an *equivalence relation*, provided it is

- reflexive, i.e.  $xRx$  for each element  $x$  in the set, and
- symmetric, i.e.  $xRy$  implies  $yRx$ , and
- transitive, i.e.  $xRy$ ,  $yRz$  implies  $xRz$ , for any elements  $x, y$  and  $z$  in the set.

For example, the relation  $R =$  'is greater than' will fulfill neither condition a), nor b). Whereas e.g.  $R =$  'is equal to' does fulfill all the three conditions, thus being an equivalence relation.

It can be proven that if  $R$  is an equivalence relation in set  $A$ , then  $R$  can be used to divide the set  $A$  into distinct subsets  $A_i$  such that each element in the set  $A$  will be contained in one and only one subset  $A_i$ . This subdivision will be such that any two elements — say  $x, y$  —

will both be contained in a common subset if and only if  $xRy$ .

Such *subsets*, consisting of 'equivalent' elements (i.e. elements connected by an equivalence relation), are called *equivalence classes*. For these basic concepts, see e.g. LIP-SCHUTZ (1964).

## App. H. Number of crosses needed (in a deterministic, dynamically optimized system) to subdivide a population into equivalence classes of incompatibility

Consider expression (6) (see 3.2.2.1), which presents the number of crosses needed ( $K$ ). Heuristically, with regard to equivalence class structure  $\bar{n}$  and a happy chance in identifying order  $\delta$ , a *minimal required number* is supposed to be

$$(7) \min_{\bar{n}, \delta} K(N, M, \bar{n}; \delta) = N + \frac{M(M-3)}{2},$$

which is attained at  $\bar{n} = (N-M+1, 1, \dots, 1)$  and  $\delta(1) = 1$ , i.e. when the greatest class will be identified first and the other classes contain but a single ramet each. Respectively, the *maximal required number* is supposed to be expressed by

$$(8) \max_{\bar{n}, \delta} K(N, M, \bar{n}; \delta) = MN - \frac{M(M+1)}{2},$$

which is attained at  $\bar{n} = (N-M+1, 1, \dots, 1)$  and  $\delta(M) = 1$ , i.e. in a similar population as above, but with the greatest class being identified last. In a population with only *one equivalence class*, both of these expressions reduce to

$$(9) K(N, 1, \bar{n}; \delta) = N - 1.$$

A general expression for the *expectance* (average value) of  $K$ , in regard to values of  $\delta$ , will be

$$(10) \bar{K}(N, M, \bar{n}) = -M + \frac{1}{N} \cdot$$

$$\sum_{\delta} \frac{\prod_{i=1}^M n_{\delta(i)}}{\prod_{i=1}^{M-1} \sum_{k=i+1}^M n_{\delta(k)}} \cdot \sum_{j=1}^M j \cdot n_{\delta(j)}.$$

Two special cases out of this expression will be presented:

$$(11) \quad \bar{K}(N, M, \bar{n}) = \frac{M(M-1)}{2} + \frac{(N-M)}{\binom{N}{M-1}}$$

$$\cdot \sum_{j=1}^M \left( \frac{N-j}{M-j} \right) \cdot j,$$

if  $\bar{n} = (N-M+1, 1, \dots, 1)$ , ie. in an *extremely unevenly classed* population, and

$$(12) \quad \bar{K}(N, M, \bar{n}) = -M + \frac{N(M+1)}{2},$$

if  $\bar{n} = (N/M, N/M, \dots, N/M)$ , ie. in an *extremely evenly classed* population.

### App. I. A single move method for finding a locally optimal hypothesis

#### I.1. One ramet changes the equivalence class

The case a) above is studied more closely. Let a single ramet, denoted by  $r_0$ , change equivalence class from  $e_i$  to  $e_j$ , with a change of a hypothesis from  $H_i$  to  $H_j$ . Then, the interpretation of all crosses remains unaltered, except the *crosses involving* the ramet  $r_0$ . Namely, all crosses of ramet  $r_0$  with any ramet in its previous class  $e_i$ , change from a 'self' to a 'compatible' type, ie. belong to a set of changed cases [is-jc]. Respectively, all crosses of ramet  $r_0$  with its new class  $e_j$ , represent a change from a 'compatible' to a 'self' type cross, ie. belong to a set [ic-js]. Hence, denoting a cross  $r \times r_0$  as well as its reciprocal  $r_0 \times r$  by an index number  $k_r$ , the number # [is-jc | l] can be expressed as

$$\# [is-jc | l] = \sum_{\substack{r \in e_i \\ r \neq r_0}} m_{k_r, l} = \sum_{\substack{r \in e_i \\ r \neq r_0}} \# [r \times r_0 \text{ or } r_0 \times r | l].$$

Respectively, # [ic-js | l] can be rewritten as

$$\# [ic-js | l] = \sum_{\substack{r \in e_j \\ r \neq r_0}} m_{k_r, l} = \sum_{\substack{r \in e_j \\ r \neq r_0}} \# [r \times r_0 \text{ or } r_0 \times r | l].$$

Furthermore,

$$\begin{aligned} m_{js}(l) &= m_{is}(l) - \# [is-jc | l] + \# [ic-js | l] \\ &= m_{is}(l) - \sum_{\substack{r \in e_i \\ r \neq r_0}} m_{k_r, l} + \sum_{\substack{r \in e_j \\ r \neq r_0}} m_{k_r, l} \end{aligned}$$

and respectively,

$$m_{jc}(l) = m_{ic}(l) - \sum_{\substack{r \in e_j \\ r \neq r_0}} m_{k_r, l} + \sum_{\substack{r \in e_i \\ r \neq r_0}} m_{k_r, l}.$$

Hence, utilizing a short notation  $m_k = \sum_l m_{k, l}$ , we can write

$$m_{js} = m_{is} - \sum_{\substack{r \in e_i \\ r \neq r_0}} m_{k_r} + \sum_{\substack{r \in e_j \\ r \neq r_0}} m_{k_r}, \text{ and}$$

$$m_{jc} = m_{ic} - \sum_{\substack{r \in e_j \\ r \neq r_0}} m_{k_r} + \sum_{\substack{r \in e_i \\ r \neq r_0}} m_{k_r}.$$

Assuming  $m_{js}(l) > m_{is}(l)$  implies  $m_{jc}(l) < m_{ic}(l)$ . Now, ejecting the symbols  $A_{ij}$  and  $A_{ij}(l)$  for the appropriate terms in the coefficient part (20), we will attain at

$$(21) \quad A_{ij}(l) \equiv \frac{m_{js}(l)!}{m_{is}(l)!} \cdot \frac{m_{jc}(l)!}{m_{ic}(l)!}$$

$$= \begin{cases} \frac{[m_{is}(l)+1] \dots m_{js}(l)}{[m_{jc}(l)+1] \dots m_{ic}(l)} & \text{if } m_{js}(l) > m_{is}(l), \text{ or} \\ 1, & \text{if } m_{js}(l) = m_{is}(l), \text{ or} \\ \frac{[m_{ic}(l)+1] \dots m_{jc}(l)}{[m_{js}(l)+1] \dots m_{is}(l)} & \text{if } m_{js}(l) < m_{is}(l). \end{cases}$$

Similarly,

$$(22) \quad A_{ij} \equiv \frac{m_{is}!}{m_{js}!} \cdot \frac{m_{ic}!}{m_{jc}!}$$

$$= \begin{cases} \frac{[m_{ic}+1] \dots m_{ic}}{[m_{is}+1] \dots m_{js}} & \text{if } m_{js} > m_{is}, \text{ or} \\ 1, & \text{if } m_{js} = m_{is}, \text{ or} \\ \frac{[m_{js}+1] \dots m_{is}}{[m_{ic}+1] \dots m_{jc}} & \text{if } m_{js} < m_{is}. \end{cases}$$

With such notations, expression (19) can be presented as follows, in a directly calculable form:

$$(23) \quad \frac{L_i}{L_j} = A_{ij} \cdot \prod_l A_{ij}(l) \cdot \prod_l \left\{ \frac{p_s(l)}{p_c(l)} \right\}^{\left[ \sum_{\substack{r \in e_i \\ r \neq r_0}} m_{k_r, l} - \sum_{\substack{r \in e_j \\ r \neq r_0}} m_{k_r, l} \right]},$$

or in short,

$$= C_{ij} \cdot T_{ij}.$$

There,  $C_{ij} \equiv A_{ij} \cdot \prod_l A_{ij}(l)$  is called the *coefficient part*, and  $T_{ij}$  the *probability part*, respectively.



### 1.2. One equivalence class disappears

Studying the case b) above, let us assume that the previous class,  $e_i$ , of the ramet  $r_0$  totally disappears, when we change from the hypothesis  $H_i$  to  $H_j$ . Then, in  $H_i$ , the equivalence class  $e_i$  necessarily contains only one ramet, that is  $e_i = [r_0]$ . Hence, the number of 'self' type crosses will at least not decrease, since  $\# [is-jc | l] = 0$ . The general formulas derived above are, however, still applicable, remembering that the sum over  $r \in e_i$ , where  $r \neq r_0$ , should now be considered to be *empty*, and always  $m_{js}(l) \geq m_{is}(l)$ . The reduced forms of the general expressions will, however, be given explicitly below, for the sake of clarity:

$$(21') \quad A_{ij}(l) \equiv \frac{m_{js}(l)!}{m_{is}(l)!} \cdot \frac{m_{jc}(l)!}{m_{ic}(l)!}$$

$$= \begin{cases} \frac{[m_{is}(l)+1] \dots m_{js}(l)}{[m_{jc}(l)+1] \dots m_{ic}(l)} \\ \text{if } m_{js}(l) > m_{is}(l), \text{ or} \\ 1, \text{ if } m_{js}(l) = m_{is}(l). \end{cases}$$

Respectively,

$$(22') \quad A_{ij} \equiv \frac{m_{is}!}{m_{js}!} \cdot \frac{m_{ic}!}{m_{jc}!}$$

$$= \begin{cases} \frac{[m_{jc}+1] \dots m_{ic}}{[m_{is}+1] \dots m_{js}}, \text{ if } m_{js} > m_{is}, \text{ or} \\ 1, \text{ if } m_{js} = m_{is}. \end{cases}$$

Finally,

$$(23') \quad \frac{L_i}{L_j} = A_{ij} \cdot \prod_l A_{ij}(l) \cdot \prod_l \left\{ \frac{p_s(l)}{p_c(l)} \right\}^{[-\sum_{\substack{r \in e_i \\ r \neq r_0}} m_{kr,l}]}$$

### 1.3. A new equivalence class appears

The third type of a one ramet move is the case c) above. Regarding a change from the hypothesis  $H_i$  into  $H_j$ , the 'previous' equivalence class,  $e_i$ , of the ramet  $r_0$  contained more than one ramet, but its 'forthcoming' class,  $e_j$ , is a newborn one and consist of  $r_0$  alone. That is,  $e_j = [r_0]$ . Therefore, any crosses  $r \times r_0$  or  $r_0 \times r$  fulfilling the condition  $r \in e_j$  and  $r \neq r_0$ , do not exist. Nor do any changes of the type  $[ic-jc | l]$  occur. Accordingly, all the sums over  $r \in e_i$ ,  $r \neq r_0$ , must now be regarded *empty*, and secondly,  $m_{js}(l)$  will never be greater than  $m_{is}(l)$ . Thus, the general expressions above can now be written into a slightly simpler form:

$$(21'') \quad A_{ij}(l) \equiv \frac{m_{js}(l)!}{m_{is}(l)!} \cdot \frac{m_{jc}(l)!}{m_{ic}(l)!}$$

$$= \begin{cases} 1, \text{ if } m_{js}(l) = m_{is}(l), \text{ or} \\ \frac{[m_{ic}(l)+1] \dots m_{jc}(l)}{[m_{js}(l)+1] \dots m_{is}(l)} \\ \text{if } m_{js}(l) < m_{is}(l). \end{cases}$$

Similarly,

$$(22'') \quad A_{ij} \equiv \frac{m_{is}!}{m_{js}!} \cdot \frac{m_{ic}!}{m_{jc}!}$$

$$= \begin{cases} 1, \text{ if } m_{js} = m_{is}, \text{ or} \\ \frac{[m_{js}+1] \dots m_{is}}{[m_{ic}+1] \dots m_{jc}}, \text{ if } m_{js} < m_{is}. \end{cases}$$

Finally,

$$(23'') \quad \frac{L_i}{L_j} = A_{ij} \cdot \prod_l A_{ij}(l) \cdot \prod_l \left\{ \frac{p_s(l)}{p_c(l)} \right\}^{[\sum_{\substack{r \in e_i \\ r \neq r_0}} m_{kr,l}]}$$

### 1.4. Short-hand version

In each cycle, there is a *heavy load of calculations*. Provided there are  $n$  ramets and  $m$  equivalence classes in  $H_i$ , out of which  $m_1$  classes are uniclinal (ie. consist of only one ramet each), then there will be

$$(24) \quad n \cdot m - m_1(m_1 + 1) / 2$$

different hypotheses  $H_j$  surrounding the "starting" hypothesis  $H_i$  at the distance of one ramet move. Accordingly many likelihood ratios (23) should be calculated in the cycle. For example, with  $H_i$  constituting a subdivision of 30 ramets into 5 equivalence classes, out of which 2 are uniclinal ones, a total of 147 surrounding hypotheses will be compared to  $H_i$  during the present cycle in the algorithm.

The most cumbersome term in calculations is the coefficient part,  $C_{ij}$ . Therefore, it was tested, if the *probability part*  $T_{ij}$  would *alone* suffice to serve as a *criterion function*.

#### 1.4.1. Behaviour of the probability part

The *behaviour* of the proposed criterion function  $T_{ij}$  is studied in a certain *simple situation*: it is applied for a choice between three hypotheses  $H_i$ ,  $H_j$  and  $H_{new}$ . It is

assumed that among the crosses made, there is only a *single cross* of the ramet  $r_0$  with any other ramet belonging to equivalence class  $e_i$  in  $H_i$ . This cross is assumed to result in  $l_i$  seeds. In  $H_j$ , this cross will be re-interpreted as a 'compatible' one. Similarly, it is assumed that there occurs only a *single cross* of the ramet  $r_0$  with the set of ramets constituting equivalence class  $e_j$  in  $H_j$ ; this cross resulted in  $l_j$  seeds. While being a 'compatible' type cross in  $H_i$ , this cross is re-interpreted in  $H_j$  as being of a 'self' type. The third hypothesis,  $H_{new}$ , states that the ramet  $r_0$  should constitute a new equivalence class.

Recalling from expression (23) that

$$(25) \quad T_{ij} \equiv \prod_l \left\{ \frac{p_s(l)}{p_c(l)} \right\}^{[\sum_{\substack{r \in e_i \\ r \neq r_0}} m_{kr,l} - \sum_{\substack{r \in e_j \\ r \neq r_0}} m_{kr,l}]},$$

and substituting the assumed values, the example yields

$$(26) \quad T_{ij} = \left\{ \frac{p_s(l_i)}{p_c(l_i)} \right\} / \left\{ \frac{p_s(l_j)}{p_c(l_j)} \right\}.$$

Similarly, from expression (23'') is obtained

$$(27) \quad T_{i_{new}} = \prod_l \left\{ \frac{p_s(l)}{p_c(l)} \right\}^{[\sum_{\substack{r \in e_i \\ r \neq r_0}} m_{kr,l}]}$$

Substituting the values of the example into this expression, will yield

$$(28) \quad T_{i_{new}} = \frac{p_s(l_i)}{p_c(l_i)}.$$

Concerning the probability distributions, it is at first assumed that  $p_s(l)/p_c(l)$  constitutes a *monotonously decreasing* sequence of numbers, and  $p_s(l)$  as well, when  $l = 0, 1, 2, \dots$ . Hence, the expression  $T_{ij} < 1$  if and only if  $l_i < l_j$  (see Fig. 10c). Thus, using  $T_{ij}$  as a criterion function,  $H_j$  is supported against  $H_i$  if and only if the cross of the 'differing' ramet  $r_0$  with equivalence class  $e_j$ , resulted in less seed than its cross with the rivalling equivalence class  $e_i$ . That is, when choosing between two rivalling equivalence classes,  $e_i$  and  $e_j$ , the ramet  $r_0$  is classified into the equivalence class, crossed with which it gave *less seed*. Such a result is in accordance with an intuitive, heuristic reasoning concerning this case. Namely, a 'self' cross will monotonously with less frequency yield more seeds. Thus, when a cross results in less seed, it is easier to believe that it may have originated from a 'self' cross than when a cross results in more seeds. In other words, for the abandoning of a 'self' explanation of an isolated cross result (when weighed against an alternative 'compatible' explanation), it is "natural" to choose a one-tailed critical region (great seed numbers).

Consider next the behaviour of the short-hand criterion function by applying more general and therefore *more realistic probability distributions*. Therefore, a single *assumption* is made: that there exists a value, say  $l_{sc}$ , such that for each  $0 \leq l \leq l_{sc}$ , the probability quotient  $p_s(l)/p_c(l)$  is at least 1, and for each  $l$  value over  $l_{sc}$ , the value of the probability quotient is less than one. The previous example, a monotonously decreasing function, fulfills these conditions, and is included as a very special case. In an incompatible plant species, a function of the kind illustrated in Fig. 10a), is however much more plausible. Both of the probability distributions,  $p_s(l)$  and  $p_c(l)$  with  $l = 0, 1, 2, \dots$ , should usually have their own mode values, the former — depending on the strength of the incompatibility — considerably near the origin, and the latter at a markedly higher seed number (see Fig. 11). In a self-incompatible plant species, only a few seeds will be produced after 'self' type pollination, while after 'compatible' pollination, a high number of seeds are produced. These modal points are likely seen in the form of the quotient function  $p_s(l)/p_c(l)$ : the former is a point of a maximum value, and the latter a point of a minimum value, as is illustrated in Fig. 10a). No such extra assumptions are, however, needed. The form of the function may also be of a more complicated nature.

In the example,  $T_{i_{new}} = p_s(l_i)/p_c(l_i) < 1$  if and only if  $l_i > l_{sc}$ . That is, if the seed set in a cross of  $r_0$  with equivalence class  $e_i$ , is high enough that such number could be more often acquired in a 'compatible' than in a 'self' type cross, then  $H_{new}$  is supported against  $H_i$ .

Sometimes also  $H_j$  is supported over  $H_i$ . Then, applying the algorithm (see above), during a cycle, the surrounding hypothesis with the strongest support against the starting hypothesis  $H_i$ , is chosen. Therefore,  $H_{new}$  is chosen instead of  $H_j$ , provided its criterion function with respect to  $H_i$  (ie.  $T_{i_{new}}$ ) is lower than that of  $H_j$  (ie.  $T_{ij}$ ). Now, substituting the values from above,

$$(29) \quad \frac{T_{ij}}{T_{i_{new}}} = 1 / \left\{ \frac{p_s(l_j)}{p_c(l_j)} \right\}.$$

Hence,  $T_{i_{new}} < T_{ij}$  if and only if  $l_j > l_{sc}$ , (irrespective of the value of  $l_i$ ).

Combining these two results from above,  $H_{new}$  is chosen if and only if both  $l_i$  and  $l_j$  exceed the value of  $l_{sc}$ ; ie. if, regarding  $r_0$ , a 'compatible' explanation is more likely than a 'self' explanation, with respect to each of the equivalence classes  $e_i$  and  $e_j$  (see Fig. 10a).

Consider next, on which conditions  $H_j$  would be chosen as a new starting hypothesis. Hypothesis  $H_j$  will be supported against  $H_i$ , ie.  $T_{ij} < 1$ , if and only if  $p_s(l_i)/p_c(l_i) < p_s(l_j)/p_c(l_j)$ . In addition,  $l_j$  must be (see above) less than  $l_{sc}$ , to avoid a choice of  $H_{new}$  instead of  $H_j$ . If  $l_i$  happens to be greater than  $l_{sc}$ , ie. falls into the "compatible" region (see Fig. 10a), then the inequality  $T_{ij} < 1$  is fulfilled and  $H_j$  is chosen. If, however, also  $l_i$  happens to be smaller than  $l_{sc}$ , interpretations are not so straightforward. For instance, if the quotient function has a mode, ie. a local maximum value in the interval  $[0, l_{sc}]$ , then even certain values  $l_i < l_j$  serve to support  $H_j$  against  $H_i$ . Namely, all  $l_i$  values not "too close" (determined by  $l_j$ ) to the maximum point (see Fig. 10b). The

situation is, however, simpler in a strictly self-incompatible plant species. Since its  $p_s(l)$  distribution plausibly has the mode at  $l = 0$  seeds, we can readily assume that in such a species, the quotient function  $p_s(l)/p_c(l)$  is monotonously decreasing in the interval  $0 \leq l \leq l_{sc}$ . Therefore, in a strictly self-incompatible plant species,  $H_j$  is chosen if and only if  $l_j < l_{sc}$  and  $l_i > l_j$  (see Fig. 10c).

Thus, in this simple example and in a strictly self-incompatible species, concerning an algorithm cycle starting with hypothesis  $H_i$ , the following decision rules have been derived for the choice between three rivaling hypotheses —  $H_i$ , and two further hypotheses at a one move distance from it, i.e.  $H_j$  and  $H_{new}$ :

$$(30) \quad \begin{cases} H_{new} \text{ will be chosen, if and only} \\ \text{if } l_i > l_{sc} \text{ and } l_j > l_{sc}, \\ H_j \text{ will be chosen, if and only} \\ \text{if } l_i > l_j \text{ and } l_j \leq l_{sc}, \text{ and} \\ H_i \text{ remains to be chosen in all} \\ \text{other cases.} \end{cases}$$

#### 1.4.2. Short-hand version is more favourable

The decision rules derived above, pertaining to the simple example, appear to be very simple. An extra inquiry is, however, required in order to determine, how powerful the short-hand method is in the selection of the right hypothesis among nearby alternatives.

As an example, let us study the choice between  $H_i$  and  $H_{new}$ . Postulating  $H_i$  holds,

$$(31) \quad \begin{aligned} P \{\text{wrong choice}\} &= P \{H_{new} \text{ is chosen}\} \\ &= P \{T_{i \text{ new}} < 1\} = P \{p_s(l_s) / p_c(l_s) < 1\} \\ &= P \{l_s > l_{sc}\}, \end{aligned}$$

which would generally be fairly small (see Fig. 11a), at least much less than one half. Similarly, postulating  $H_{new}$  holds, it can be determined that

$$(32) \quad \begin{aligned} P \{\text{wrong choice}\} &= P \{H_i \text{ is chosen}\} \\ &= P \{T_{i \text{ new}} > 1\} = P \{p_s(l_c) / p_c(l_c) > 1\} \\ &= P \{l_c < l_{sc}\}. \end{aligned}$$

Though also being a rare occasion, this latter type of error would as a rule occur somewhat more frequently than the former one (Fig. 11b).

Using the short-hand version of the method, correct classifications would thus be much more frequent than each type of the erroneous ones. In other words: at least in a simple situation, the short-hand version manages well, performing the right classification in a powerful way.

For a sake of a comparison, consider next the power of the 'general', unshortened criterion function. Substituting the values from the example into expressions (21'') and (22''), we will obtain the coefficient part

$$(33) \quad C_{i \text{ new}} = \frac{m_{is}}{[m_{ic} + 1]} \cdot \frac{[m_{ic}(l_i) + 1]}{m_{is}(l_i)}.$$

Therefore,

$$(34) \quad \begin{aligned} \frac{L_i}{L_{new}} &\equiv C_{i \text{ new}} \cdot T_{i \text{ new}} \\ &= \left( \frac{[m_{ic}(l_i) + 1]}{[m_{ic} + 1]} / \frac{m_{is}(l_i)}{m_{is}} \right) \cdot \frac{p_s(l_i)}{p_c(l_i)}. \end{aligned}$$

Postulating that  $H_i$  holds — and remembering that at least one 'self' type cross resulting in  $l_i$  drupelets was assumed as having occurred (i.e.  $m_u(l_i) \geq 1$ ) — the following expression is obtained

$$(35) \quad \begin{aligned} P \{\text{wrong choice}\} &= P \left\{ \frac{L_i}{L_{new}} < 1 \mid m_{is}(l_i) \geq 1 \right\} \\ &= P \left\{ \left( \frac{[m_{ic}(l_i) + 1]}{[m_{ic} + 1]} / \frac{m_{is}(l_i)}{m_{is}} \right) \cdot \frac{p_s(l_i)}{p_c(l_i)} \right. \\ &\quad \left. < 1 \mid m_{is}(l_i) \geq 1 \right\}. \end{aligned}$$

In this case, it can be expressed as

$$(36) \quad \begin{aligned} P \left\{ \frac{L_i}{L_{new}} < 1 \mid m_{is}(l_i) \geq 1 \right\} &= \\ &= \frac{\sum_k p_s(k) \cdot P \left\{ \left( \frac{m_{ic}(k) + 1}{m_{ic} + 1} / \frac{m_{is}(k)}{m_{is}} \right) \cdot \frac{p_s(k)}{p_c(k)} < 1 \right\}}{\sum_k p_s(k) \cdot P \{m_{is}(k) \geq 1\}}. \end{aligned}$$

After some steps this yields

$$(37) \quad \begin{aligned} P \{\text{wrong choice}\} &= P \{H_{new} \text{ is chosen}\} \\ &= \frac{\sum_k p_s(k) \cdot \sum_{z=1}^{m_{is}} \binom{m_{is}}{z} \cdot p_s(k)^z [1 - p_s(k)]^{m_{is}-z}}{1 - \sum_k p_s(k) \cdot [1 - p_s(k)]^{m_{is}}} \\ &\quad \cdot \sum_{y=0}^{\min \{A_z, m_{ic}\}} \binom{m_{ic}}{y} \cdot p_c(k)^y [1 - p_c(k)]^{m_{ic}-y} \end{aligned}$$

where  $A_z$  denotes the greatest whole number that is less than

$$\frac{(m_{ic} + 1) \cdot p_c(k)}{m_{is} \cdot p_s(k)} \cdot z - 1,$$

and where the value of the respective sum is defined to be zero, if  $A_z < 0$ .

This is a complicated formula, the behaviour of which is difficult to study in general, without detailed knowledge concerning the probability distributions and values of  $m_{ic}$  as well as  $m_{is}$  realized in the crossing experiment.

The *expected value* of the 'general' criterion function can be shown to be

$$(38) \quad E \left\{ \frac{L_i}{L_{new}} \mid m_{is}(l_s) \geq 1 \right\} \\ = \frac{m_{is}}{m_{ic} + 1} \cdot \sum_k \frac{p_s(k)^2}{p_c(k)} \cdot \frac{[p_c(k) \cdot m_{ic} + 1]}{1 - [1 - p_s(k)]^{m_{is}}} \\ \cdot \sum_{l=1}^{m_{is}} \frac{1}{l} \cdot \binom{m_{is}}{l} \cdot p_s(k)^l [1 - p_s(k)]^{m_{is}-l}.$$

Also this formula is too complicated to give a decisive general answer about the classificatory power of the 'general' method.

Some hints can be acquired by studying the *limit behaviour* of the criterion function, however. Expression (36) can be rewritten as

$$(39) \quad P \left\{ \frac{L_i}{L_{new}} < 1 \mid m_{is}(l_i) \geq 1 \right\} \\ = P \left\{ \frac{\hat{p}_c(l_i)}{\hat{p}_s(l_i)} \cdot \frac{p_s(l_i)}{p_c(l_i)} < 1 \mid m_{is}(l_i) \geq 1 \right\}.$$

Here the estimate of  $p_c(l_i)$  is, however, biased upwards. When the *number of crosses* in the study is *increased*, this bias will diminish without a limit, so that the estimate in question will converge towards the respective probability value  $p_c(l_i)$ . Due to the condition that zero values of  $m_{is}(l_i)$  are excluded, also the estimate of  $p_s(l_i)$  is biased upwards. This bias will also diminish with increasing number of crosses. Therefore, the *large sample* form of the error probability is

$$(40) \quad P \{\text{wrong choice}\} = P \{H_{new} \text{ is chosen}\} \\ = P \left\{ \frac{\hat{p}_s(l_i)}{\hat{p}_c(l_i)} > \frac{p_s(l_i)}{p_c(l_i)} \right\},$$

where both estimated values are almost unbiased.

This result is not promising but rather suggests that, using the 'general' version of the criterion function, *wrong decisions* would be made *quite often*. This should pertain at least to the later cycles of crossing, when a considerable number of crosses have already been made. In the very beginning, however, the bias in the estimate of

$p_c(l_i)$  will be greater, and usually not wholly compensated by the opposing bias in  $p_s(l_i)$ . Accordingly, wrong classifications are likely to be less common then.

In conclusion, the classificatory power of the 'general' method remains *without confirmation*. In addition, its power exhibits *unfavourable variation* between studies — due to chance, as well as in the course of a study also to systematic effects.

On the contrary, at least in a simple example, the 'short-hand' criterion function displays a clear-cut, *constant and high classificatory power*.

Hence there are some indications that the 'general' method is less satisfactory than its 'short-hand' version. If this fact is accepted, then this result may be explained through the following proposal. Reconsidering expression (17), it can be suggested that in the 'general' method, there is a *misinterpretation* concerning what is to be regarded as an 'actual result' in the experiment. The actual result was understood to be *any* series of crosses yielding the required total numbers of *l*-drupelet berries, in 'self' type and respectively in 'compatible' crosses (typed *according to the hypothesis*). Thus, in the 'general' method, not only the really recorded (actualized) pairwise cross results are considered, but in addition a *large number of non-actualized*, purely ideal (possible) cross results. For instance, if the recorded result in a 'compatible' cross  $r_{k1} \times r_{k2}$  was  $l_{12}$ , and in another 'compatible' cross  $r_{k3} \times r_{k4}$  the recorded result was  $l_{34}$ , in the 'general' method, also the *non-occurred* result  $l_{34}$  in  $r_{k1} \times r_{k2}$  and  $l_{12}$  in  $r_{k3} \times r_{k4}$ , would in addition be counted in the 'actual result'. Since for practical calculations, the original cross results matrix  $\hat{M}$  was *pooled* into the two vectors  $\bar{m}_{is}$  and  $\bar{m}_{ic}$  (see 3.3.2.), such a confusion became possible. These pooled data should however *not* be considered as an *actual result* in *this* study, since they have *lost information* essential for that purpose!

Hence, the multinomial type coefficient part in the 'general' method has arisen erroneously, due to a mistake of including an array of "ghost" results. Perhaps situations exist, where such an interpretation of data might be well based. However in the present study, the 'short-hand' version can not be regarded as a modification. On the contrary, it will be regarded as the *correct form of the method*.

## App. J. Average efficiency of procedures I and II in the selection of a proper ramet as the first crossing partner

In order to compare the *average efficiency* of procedures I and II (see 3.3.5.2, Theorem II), the respective expectances (denoting them by  $E_I$  and  $E_{II}$ ) of the increase in support of  $H_0^{(n)}$  against its "r<sub>1</sub>, r<sub>2</sub>-opposite" hypotheses, under each procedure, are calculated. Applying the Total Probability Theorem, it is straightforward to obtain the formula

$$(52) \quad E_I \{ \delta^{(n)} \log T_{0j} \} = \text{incr} - P_I \{ r_{1, r_2} \text{-false} \} \cdot [\text{maxincr} + \text{incr}].$$

Here *incr* and *maxincr* are the expectances of the increases described in Theorem I, ie.

$$(53) \text{ incr} = E \left\{ \log \left[ \frac{p_s(l_s)}{p_c(l_s)} \right] \right\}, \text{ and}$$

$$\text{maxincr} = E \left\{ -\log \left[ \frac{p_s(l_c)}{p_c(l_c)} \right] \right\}.$$

It is natural to assume both *incr* and *maxincr* to be positive. Though, in this study a low seed set indicates a 'self' type cross fairly less decisively than a high seed set indicates a 'compatible' cross (see above). Therefore, the *incr* will most likely not be high in value.

It can be readily observed that under *procedure I*, on average an actual decrease in support of  $H_0^{(n)}$  against its " $r_1, r_2$ -opposite" hypotheses is achieved *if and only if*

$$(54) P_I \{r_1, r_2 \text{ -false}\} > \left( \frac{\text{incr}}{\text{maxincr} + \text{incr}} \right).$$

Similarly,

$$(55) E_{II} \{ \delta^{(n)} \log T_{0j} \} \\ = \text{maxincr} - P_{II} \{r_1, r_2 \text{ -false}\} \cdot \\ \cdot [\text{maxincr} + \text{incr}],$$

and therefore, under *procedure II*, an average decrease in support of  $H_0^{(n)}$  against its " $r_1, r_2$ -opposite" hypotheses is achieved *if and only if*

$$(56) P_{II} \{r_1, r_2 \text{ -false}\} > \left( \frac{\text{maxincr}}{\text{maxincr} + \text{incr}} \right).$$

Furthermore, the support in question will on average decrease more under *I* than under *II*, if and only if

$$(57) P_{II} \{r_1, r_2 \text{ -false}\} - P_I \{r_1, r_2 \text{ -false}\} \\ < \frac{\text{maxincr} - \text{incr}}{\text{maxincr} + \text{incr}}.$$

In a pair of simple examples, it can be shown that the relative efficiency of the two procedures in replacing a false  $H_0^{(n)}$ , depends on the real subdivision of the particular population into equivalence classes.

*Example 1.* Suppose that in reality, all ramets belong to a single equivalence class, while our  $H_0^{(n)}$  states that there are  $m_0$  equivalence classes. Then, if we choose whatsoever pair of ramets from different classes according to  $H_0^{(n)}$ , their stated relation in  $H_0^{(n)}$  will be false, that is  $P_{II} \{r_1, r_2 \text{ -false}\} = 1$ . Conversely, choosing any ramet pair within an equivalence class according to  $H_0^{(n)}$ , will never hit at a false relation, ie.  $P_I \{r_1, r_2 \text{ -false}\} = 0$ . Therefore, the condition (57) will never be fulfilled, and thus *procedure II* will be more efficient. Since the condi-

tion (56) is fulfilled while (54) is not, *procedure II* will actually decrease the support as desired, while *procedure I* would increase it instead.

*Example 2.* Suppose that in reality, each ramet constitutes its own equivalence class, ie. all classes are uniclonal. As before,  $H_0^{(n)}$  states that there are  $m_0$  equivalence classes. Then, if we choose whatsoever pair of ramets from different classes according to  $H_0^{(n)}$ , their relation in  $H_0^{(n)}$  will be true, ie.  $P_{II} \{r_1, r_2 \text{ -false}\} = 0$ . On the contrary, choosing any ramet pair within an equivalence class according to  $H_0^{(n)}$ , will always direct the selection at a false relation, ie.  $P_I \{r_1, r_2 \text{ -false}\} = 1$ . Therefore, the condition (57) will always be fulfilled, and thus *procedure I* is more efficient. Since the condition (54) is fulfilled while (56) is not, *procedure I* will actually decrease the support as desired, while *procedure II* would increase it.

These examples show that, if we knew the population to consist actually of only one class, the most efficient way to complete the classification, would be to follow the *procedure II*, that is to choose the two ramets to be crossed from different equivalence classes (according to our false  $H_0^{(n)}$ ) each. Since we do not have such a knowledge, however, relying in all cases on the *procedure I*, may still prove to be more desirable. After all, the tedious cases requiring prolonged crossing, will be the populations with many equivalence classes, while in populations with a single class, we can also manage by relying upon a design that is less than the most strongly optimized. Towards the end of the crossing study, bearing in mind the possibility of an uniclassical population, some crossing between classes might also be advisable. This would facilitate the amalgamation of classes occasionally kept apart falsely, eg. due to an erroneous or unluckily extreme cross result.

In the previous, extreme examples, one of the "hitting" probabilities, either  $P_I \{r_1, r_2 \text{ -false}\}$  or  $P_{II} \{r_1, r_2 \text{ -false}\}$ , was maximal in value. In other cases, however — disregarding whether the choice is made within a class or between two classes — a random sampling of partners under a procedure, may yield a fortunate hit at a false relation only rarely. Hence, especially towards the end of the study, when there will be only a small number of false relations left in  $H_0^{(n)}$ , the most pertinent problem is, how to succeed in hitting at a false relation at all, no more a (minor) choice between the two procedures. Namely, if the "hitting" probability is low enough, each of the procedures will become incapable of decreasing on average the support of  $H_0^{(n)}$ . This is obvious when the following example is considered.

*Example 3.* Suppose that in reality (ie. according to the true hypothesis  $H_1$ ), there are  $m_1$  equivalence classes in the population. Let  $H_0^{(n)}$  be already as near to the true hypothesis as at a one ramet move distance to it. Let in  $H_0^{(n)}$  there be  $m_0 = m_1$  equivalence classes, of sizes  $n_1, n_2, \dots, n_{m_0}$ . Suppose that in  $H_0^{(n)}$ , the ramet  $r_1 \in e_1$ , while in  $H_1$ ,  $r_1 \in e_2$ . Since the number of equivalence classes is assumed to remain unchanged,  $n_1 > 1$  and  $n_2 > 1$  will be implied.

Then, under *procedure I*, there will be  $\sum_i n_i(n_i - 1)/2$  different ways choosing a pair of ramets within a common class in  $H_0^{(n)}$ . Out of these,  $n_1 - 1$  will hit at a false relation in  $H_0^{(n)}$  (ie. at any relation of  $r_1$  with other ramets in  $e_1$ ). Hence,

$$P_1 [r_1, r_2\text{-false}] = \frac{(n_1 - 1)}{\frac{1}{2}[\sum_i n_i^2 - N]}$$

where N denotes the total number of ramets. Similarly, choosing the ramets  $r_1$  and  $r_2$  from two *different* equivalence classes in  $H_0^{(n)}$ , there will be  $n_2$  different possible hits at a false relation, namely those of  $r_1$  with any ramet in  $e_2$  (according to  $H_0^{(n)}$ ). In total, there will be  $\sum n_i(N - n_i)/2$  such between-class choices possible in  $H_0^{(n)}$ , and therefore

$$P_{II} [r_1, r_2\text{-false}] = \frac{n_2}{\frac{1}{2}[N^2 - \sum_i n_i^2]}$$

In a special case with  $m_0$  equally frequent classes in the population, these formulas will reduce to

$$P_1 [r_1, r_2\text{-false}] = \frac{2}{N}, \text{ and}$$

$$P_{II} [r_1, r_2\text{-false}] = \frac{2}{N(m_0 - 1)}$$

In such a special population, condition (57) will always be satisfied, provided the population contains at least two equivalence classes. Thence, *procedure I* would actually be preferred over *procedure II*. Depending on the properties of the particular quotient function  $p_c(l)/p_e(l)$ , even *procedure I* may, however, prove *inefficient*, in the sense that it does not any longer yield an average decrease in support of  $H_0^{(n)}$ . Since, to fulfill the condition (54),  $\text{max-incr}$  should be greater than  $[(N-2)/2] \cdot \text{incr}$ . Because the values under consideration are logarithmic ones (see 53), such a requirement may be hard to satisfy but by an "excellent" quotient function (see Figs 10 and 11).

### App. K. Low similarity to its 'own' class is less common in a greater subclass

The following conditional probability (see 3.3.5.4) is studied

$$(63) \quad P \{ \text{ramet } r_{11}(1) \text{ exhibits a "low similarity" to its 'own' class } e_1 \mid \text{classification } H_0^{(n)} \text{ has occurred} \} \\ = P \{ S^{(n)}(r_{11}(1), e_1) < z \mid \text{classification } H_0^{(n)} \} \\ = \frac{P \{ S^{(n)}(r_{11}(1), e_1) < z \text{ and classification } H_0^{(n)} \}}{P \{ \text{classification } H_0^{(n)} \}}$$

First we will evaluate the denominator. For the considered classification to occur, the similarity of any ramet

$$(63') \quad P \{ S^{(n)}(r_{11}(1), e_1) < z \mid \text{classif. } H_0^{(n)} \} \\ \approx \frac{P \{ S^{(n)}(r_{11}(1), e_v) \leq S^{(n)}(r_{11}(1), e_1) < z \text{ for } v = 2, \dots, m_0 + 1 \}}{P \{ S^{(n)}(r_{11}(1), e_v) \leq S^{(n)}(r_{11}(1), e_1) \text{ for } v = 2, \dots, m_0 + 1 \}} \\ = \frac{P \{ \max_{v=2, \dots, m_0+1} S^{(n)}(r_{11}(1), e_v) \leq S^{(n)}(r_{11}(1), e_1) < z \}}{P \{ \max_{v=2, \dots, m_0+1} S^{(n)}(r_{11}(1), e_v) \leq S^{(n)}(r_{11}(1), e_1) \}}$$

to its own class in the classification, must have been as great or greater than its similarity to any of the present other classes. To cope with classes which contain only a single ramet, we formally define  $S(r_u(1), e_u) = 0$  if  $n_u = 1$ . The class no.  $m_0 + 1$  is not actually existing in  $H_0^{(n)}$  but represents an "emerging" class, ie. the possibility of a ramet to not belong to any of the classes in  $H_0^{(n)}$ . Referring to expressions (41") and (43"), we can formally define  $S(r_u, e_u) = 0$  for  $u = m_0 + 1$ .

Hence,

$$(64) \quad P \{ \text{classification } H_0^{(n)} \} \\ = P \{ S^{(n)}(r_u(j), e_u) \geq S^{(n)}(r_u(j), e_v) \text{ for each } v = 1, \dots, m_0 + 1, v \neq u, \text{ and for each } u = 1, \dots, m_0, j = 1, \dots, n_u \}$$

Unfortunately, these similarities will not be totally independent from each other. Regarding two arbitrary ramets, say  $r_{u1}(j_1)$  and  $r_{u2}(j_2)$ , the two similarities,  $S^{(n)}(r_{u1}(j_1), e_{u1})$  and  $S^{(n)}(r_{u2}(j_2), e_{u1})$  will be interdependent, while their other similarities will not. This interdependence will arise, however, only via the crosses between just these two ramets, ie. only via one out of the  $n-1$  different crosses available to each. Considering all such similarities by different ramets to be independent from each other, will therefore not introduce a severe error, and we can obtain an approximate result

$$(65) \quad P \{ \text{classification } H_0^{(n)} \} \\ \approx \prod_{\substack{u=1, \dots, m_0 \\ j=1, \dots, n_u}} P \{ S^{(n)}(r_u(j), e_u) \geq S^{(n)}(r_u(j), e_v) \} \\ \text{for } v = 1, \dots, m_0 + 1, v \neq u \}$$

Proceeding similarly, we can obtain for the numerator

$$(66) \quad P \{ S^{(n)}(r_{11}(1), e_1) < z \text{ and classif. } H_0^{(n)} \} \\ \approx P \{ S^{(n)}(r_{11}(1), e_v) \leq S^{(n)}(r_{11}(1), e_1) < z \text{ for } v = 2, \dots, m_0 + 1 \} \\ \cdot \prod_{\substack{u=1, \dots, m_0 \\ i=1, \dots, m_i \\ n_{ui} \geq 1 \\ j=1, \dots, n_{ui} \\ u, i, j \neq 1, 1}} P \{ S^{(n)}(r_{ui}(j), e_u) \leq S^{(n)}(r_{ui}(j), e_u) \} \\ \text{for } v = 1, \dots, m_0 + 1, v \neq u \}$$

Finally, we attain at the conditional probability

Referring to expression (59), the probability distribution of  $S^{(n)}(r_{11}(1), e_v)$  is very complicated, since it is a sum of varying numbers of two different random variables. Furthermore, each of the random variables is a tedious function of a primary random variable, either  $l_s$  or  $l_c$ . Thence, the probability distribution of the *maximum* of several different  $S^{(n)}(r_{11}(1), e_v)$  items, would be nice indeed — best called a “rat’s nest” (quoting systems analytical terminology).

Therefore, aiming at utilizing the general result (63’), we must get it into a more easily applicable form. Referring to expression (59),  $S^{(n)}(r_{11}(1), e_v)$  is a sum of two sums, say  $\text{Sum}_I$  and  $\text{Sum}_{II}$ . Notice that each term in  $\text{Sum}_I$  — that is each  $\log [p_s(l_s)/p_c(l_s)]$  — is an independent random variable with identical distribution. Therefore, on the basis of the Central Limit Theorem of statistics,  $\text{Sum}_I$  will be *asymptotically normally distributed*. Denoting the number of  $\log [p_s(l_s)/p_c(l_s)]$  terms in the sum by  $C_{vs}$ , and the mean and the variance of  $\log [p_s(l_s)/p_c(l_s)]$  by  $\mu_s (>0)$  and  $\sigma_s^2$ , respectively, the normal distribution in question is

$$(67) \quad \text{Sum}_I \equiv \sum_{i=1}^{C_{vs}} \log \frac{p_s(l_s)}{p_c(l_s)} \div N(C_{vs} \cdot \mu_s, \sigma_s \sqrt{C_{vs}}).$$

Similarly,

$$(68) \quad \text{Sum}_{II} \equiv \sum_{i=1}^{C_{vc}} \log \frac{p_s(l_c)}{p_c(l_c)} \div N(C_{vc} \cdot \mu_c, \sigma_c \sqrt{C_{vc}}),$$

where  $\mu_c \leq 0$  and  $\sigma_c^2 > \sigma_s^2$ . Since  $\text{Sum}_I$  and  $\text{Sum}_{II}$  are independent, their sum will also be normally distributed, and we obtain

$$(69) \quad S^{(n)}(r_{11}(1), e_v) \div N(C_{vs} \cdot \mu_s + C_{vc} \cdot \mu_c, \sqrt{C_{vs} \cdot \sigma_s^2 + C_{vc} \cdot \sigma_c^2}).$$

Though the required distribution is now available, the conditional probability (63’) still remains unduly complicated, due to the lower boundary being a maximum value of several different random variables. Therefore, in order to find out the rough behaviour of the conditional probability, the expression is simplified by substituting a zero for the lower boundary. It represents the smallest possible value, since  $S^{(n)}(r_{11}(1), e_v) = 0$  for  $v = m_0 + 1$  (see above). It will also be a value commonly occurring as the maximum value, at least in the later stages of the crossing study, when  $H_0^{(n)}$  approaches the true hypothesis. Namely, in the later stages, a ramet  $r$  will only rarely exhibit a positive similarity value to more than a single class, since this would as a rule premise that a clear majority of the crosses of the ramet with each such “r-similar” class in question would have been ‘self’ type ones (see Corollary I). Hence, the *simplified problem* is to study the behaviour of the function

$$(63'') \quad g(C_{1s}) \equiv \frac{P\{0 \leq S^{(n)}(r_{11}(1), e_1) < z\}}{P\{0 \leq S^{(n)}(r_{11}(1), e_1)\}}.$$

In the following, we show that (at least if  $C_{1s} \cdot \mu_s + C_{1c} \cdot \mu_c > 0$ , ie. for a sufficiently large subclass)  $g(C_{1s})$  is a decreasing function of  $C_{1s}$ . That is, a *proof* is given for *Theorem III* (see 3.3.5.4).

*Proof.*

The numerator of  $g(C_{1s})$  is

$$\Phi\left(\frac{z - C_{1s} \cdot \mu_s - C_{1c} \cdot \mu_c}{\sqrt{C_{1s} \cdot \sigma_s^2 + C_{1c} \cdot \sigma_c^2}}\right) - \Phi\left(\frac{-C_{1s} \cdot \mu_s - C_{1c} \cdot \mu_c}{\sqrt{C_{1s} \cdot \sigma_s^2 + C_{1c} \cdot \sigma_c^2}}\right),$$

and the denominator is

$$1 - \Phi\left(\frac{-C_{1s} \cdot \mu_s - C_{1c} \cdot \mu_c}{\sqrt{C_{1s} \cdot \sigma_s^2 + C_{1c} \cdot \sigma_c^2}}\right),$$

where  $\Phi(x)$  denotes the cumulative distribution function of the standardized normal distribution  $N(0,1)$ . A short notation for the previous expressions is introduced, so that

$$(70) \quad g(C_{1s}) = [\Phi((z)) - \Phi((0))] / [1 - \Phi((0))].$$

The behaviour of  $g(C_{1s})$ , as a function of  $C_{1s}$ , is studied by taking its derivative. The sign of this derivative will obey the sign of the numerator of the derivative, ie. of the expression

$$[1 - \Phi((0))] \cdot [f((z)) \cdot \frac{d((z))}{d(C_{1s})} - f((0)) \cdot \frac{d((0))}{d(C_{1s})}] - [\Phi((z)) - \Phi((0))] \cdot [-f((0)) \cdot \frac{d((0))}{d(C_{1s})}].$$

That is, the derivative of  $g(C_{1s})$  will be negative *if and only if*

$$(71) \quad [1 - \Phi((0))] \cdot f((z)) \cdot \frac{d((z))}{d(C_{1s})} - [1 - \Phi((z))] \cdot f((0)) \cdot \frac{d((0))}{d(C_{1s})} < 0,$$

where  $f(x)$  denotes the probability density function of the standardized normal distribution  $N(0,1)$ . Since  $z$  is positive,  $1 - \Phi((0)) > 1 - \Phi((z))$ . The derivative

$$\frac{d((z))}{d(C_{1s})} = \frac{[-\frac{1}{2}\mu_s\sigma_s^2 \cdot C_{1s} - (\mu_s\sigma_c^2 - \frac{1}{2}\mu_c\sigma_s^2) \cdot C_{1c} - \frac{1}{2}\sigma_s^2 \cdot z]}{(\sqrt{C_{1s} \cdot \sigma_s^2 + C_{1c} \cdot \sigma_c^2})^3} < 0,$$

since  $\mu_c < 0$  and  $\mu_s > 0$ . The same applies for  $d((0))/d(C_{1s})$ , which thus is always negative, however greater (smaller in absolute value) than  $d((z))/d(C_{1s})$ . Therefore, while the truthfulness of the condition (71) may

be hard to study in general, there is a special situation, in which it can be *easily proven* to be fulfilled for certain, that is in cases where  $f(z) \geq f(0)$  applies. Remembering the notation (0) =

$$[-C_{1s} \cdot \mu_s - C_{1c} \cdot \mu_c] / [\sqrt{C_{1s} \cdot \sigma_s^2 + C_{1c} \cdot \sigma_c^2}],$$

the argument (0) is seen to be negative if and only if its numerator is negative. Therefore, by choosing  $z$  suitably small though positive,  $f(0)$  will always be smaller than  $f(z)$ , if  $C_{1s} \cdot \mu_s + C_{1c} \cdot \mu_c > 0$ . Since, then the argument (0) will be on the left tail of the standard normal density function, where the normal density is monotonously increasing.

Very plausibly, Theorem III will as a rule hold also for relatively *smaller* subclasses, ie. even if  $f(z)$  were smaller than  $f(0)$  — though not too much smaller. To acquire a stronger proof than the one given above, much more exhaustive formulas would be needed. This would be much more difficult, and more exact knowledge concerning eg. the quotient function  $p_s(t)/p_c(t)$ , would likely be needed. Thence, any attempt at a more general proof is bypassed here.

#### App. L. Probability of $n$ seeds (taken from a population with three alleles at the incompatibility locus) being of a common incompatibility genotype

Let there be three incompatibility alleles,  $s_1$ ,  $s_2$  and  $s_3$  represented in the population. Hence, the occurring genotypes will be  $s_1s_2$ ,  $s_1s_3$  and  $s_2s_3$ . Assume that their respec-

tive shares of the seeds produced in the population will be  $r_1$ ,  $r_2$  and  $r_3$ . The proportions of the different incompatibility genotypes among the seeds in the population, will then be

$$\begin{aligned} & \frac{1}{2}(r_2 + r_3) \text{ for } s_1s_2 \text{ seeds,} \\ & \frac{1}{2}(r_1 + r_3) \text{ for } s_2s_3 \text{ seeds, and} \\ & \frac{1}{2}(r_1 + r_2) \text{ for } s_1s_3 \text{ seeds.} \end{aligned}$$

If we take  $n$  seeds at random, the probability  $P_n$  that each of them will carry a common incompatibility genotype will be

$$(91) \quad P_n = \frac{1}{2^n} \cdot [(r_2 + r_3)^n + (r_1 + r_3)^n + (r_1 + r_2)^n].$$

In order to find the extreme points,  $1 - r_1 - r_2$  is substituted for  $r_3$ , and partial derivatives with respect to  $r_i$  ( $i = 1, 2$ ) are taken. There is only a single real zero point for the group of the two partial derivative equations. This root can be shown to constitute the *minimum* point of  $P_n$ . The minimum will be attained by allowing the seed shares of different incompatibility genotypes to be *equal* ( $r_1 = r_2 = r_3 = 1/3$ ). Therefore, the *maximum* value of this probability will be found somewhere at the *border* of the definition area ( $0 \leq r_i \leq 1$  for each  $i$ ,  $\sum r_i = 1$ ). Actually the maximum is attained by allowing no genotype except one, say genotype number  $k$ , to produce any seeds ( $r_k = 1$ , while  $r_i = 0$  for each  $i \neq k$ ). By substituting these values into expression (91), we will finally arrive at

$$(92) \quad \min P_n = \frac{1}{3^{n-1}} \text{ and } \max P_n = \frac{1}{2^{n-1}}.$$