3.4 Factors influencing the occurrence of tansy and aphids

For the island survey, I chose all 215 islands within an area of approximately 4400m x 5000m (Figure 4). This area included the complete inner archipelago zone. Rocks protruding on the water surface with an area of less than 0.1ha and which have no vegetation and are often overflowed by the surrounding water, were not included in the survey. The study area did not comprise the islands from the outermost archipelago zone (sea zone, Häyrén, 1914) because these islands were far away from each other, it was not possible to access them by a small motorboat.

In the survey area, all islands were checked for the presence of tansy in June and July 2001. If tansy occurred, the approximate position of the tansy groups were marked on a map or on an aerial photograph. All islands with tansy were revisited between August 10th and August 31st 2001. At this time, dispersal by winged unsexual individuals had been largely completed and the first sexual morphs had appeared (see chapter 4.4.3). Genets where sexual individuals were observed, were likely to be the sites where eggs were laid. Islands with tansy were visited once, and all tansy genets were inspected for the presence of the aphids (see below). A number of variables were measured to characterise each genet and the island itself.

3.4.1 Variables recorded for surveyed genets

Variables that were suspected to influence the presence of *M. fuscoviride* and *M. tanacetaria* on the genets were measured, as well as the occupancy of the aphids. For each tansy genet, the following nine variables were recorded:

(1) **Number of ramets** – the total number of ramets per genet was counted. In some meadows, tansy genets grew at a very close interval resulting presumably in an intermixing of ramets of different genets. Because in such cases, it was not possible to delineate a particular genet, I assigned all ramets to the same genet when they grew less than 30cm apart from the next ramet. This rule of thumb is based on the assumption that for aphids walking from one ramet to the next, it is the distance to the next ramet rather than the genetic identity that matters for successful dispersal.

(2) **Number of ramets occupied by *M. fuscoviride* and/or *M. tanacetaria*** – the number of ramets with *M. fuscoviride* and/or *M. tanacetaria*. Aphids feeding on the same ramet were defined as a colony.

(3) **Phenology** –the phenological state of every ramet was inspected and the most common phenological state was noted to represent the phenological state of the genet. At the

time of the flowering, the growth is completed. The phenological states and the time of flowering depend on the climatic conditions (Freise, 1997). The following states were distinguished: Shoot – no bud had yet been formed. The meristem was clearly seen (Figure 5). Bud –a bud was clearly visible (Figure 5). Flower – the bud had opened and the yellow flowers appeared (Figure 5). Withered – flowers had a brownish colour indicating the beginning of the seed production (Figure 5). Seed –seeds were present in the flower heads. In addition, the following states of ramets were defined: Crippled – deformed ramets which were not able to produce seeds, probably due to herbivores, drought or bad soil quality. Dead –after having set seed or because of drought, the above-ground part of the genet dried up (Figure 5).

Figure 5: The different phenological states of a tansy ramet. Pictures from the seed setting state and the crippled deformation are missing.

(4) **Maximum height** – the height of the highest ramet of a tansy genet was measured. It was assumed that the highest ramet was the most likely to get colonised by dispersing winged aphids. Tansy height depends on the water availability and therefore also on the water storage capacity of the soil type (Geisler, 1988). In August, most tansy flowered and the main growth period was already completed (Freise, 1997).

(5) **Habitat type**– tansy genets grew in different habitat types, which varied in their plant composition, the height of the vegetation, the amount of accumulated soil, the water and nutrient availability. The habitat type was likely to influence the quality of the tansy and probably the likelihood of colonisation by a migrating aphid. I distinguished the following types: Forest – *Pinus silvestris* is the most common tree in the forests on the islands. The forest soil is largely covered by lichens and mosses. Tansy grew at the edge of the forest, close to the sea shore, but was never found in the middle of the pine forest (Härri, pers. observation). Rock crevice – small crevices between rocks where some soil had accumulated. Other plants associated with tansy in this habitat type were *Solidago virgaurea, Allium schoenoprasum* and *Rumex acetosella* (Raatikainen et al, 1977, Figure 5). Rock groove – strips of vegetation in < 50cm broad crevices containing a thin soil layer (usually less than 10cm, Pokki, 1981). Because of the thin soil layer, this habitat type was very sensitive to summer droughts. Apart from tansy, the rock grooves were often colonised by *Sedum acre, Allium schoenoprasum, Festuca rubra, Solidago virgaurea, Rumex acetosella* and *Tripleurospermum maritimum* (Figure 6). Short meadow – a meadow where the vegetation was less than 50cm high. The soil layer of these short meadows was very thin and therefore this habitat type was also sensitive to droughts. Short meadows were often found in area of high wind exposure. The wind decelerated the accumulation of soil and prevented plants such as *Filipendula ulmaria* that normally reach greater heights, from exceeding 50cm. Typical representatives of this habitat were *Vicia cracca, Achillea sp., Filipendula ulmaria, Linaria vulgaris, Veronica spicata, Veronica longifolia, Sedum telephium, Allium schoenoprasum, Rumex sp., Leucanthemum sp., Festuca rubra* and *Agrostis tenuis* (Figure 6). Dry stony meadow – dry stony meadows were covered with medium-sized stones (Ø <1m). Between the stones, mostly single tansy genets were growing. Another common plant in this habitat was *Solidago virgaurea.* Stony meadow –the only difference to the dry stony meadow was the occasional flooding. *Phragmites communis, Elymys arenaria* and *Calamagrostis sp.* were mainly associated with this habitat type (Figure 6). Tall meadow – a meadow with plants mostly higher than 50cm. Plants associated with tansy in such habitats were *Filipendula ulmaria, Rubus ideaus, Carum carvi, Valeriana officinalis, Lysimachia vulgaris, Galeopsis*

sp., Rosa rugosa, Vicia cracca and *Phalaris arundinacea* (Figure 6). Wet tall meadow – in deeper depressions, the accumulation of peat has resulted in small bog-like communities characterised by *Carex canescens, Potentilla palustris, Agrostis stolonifera, Lysimachia vulgaris, Lythrum salicaria* and *Phragmites sp.* (Pokki, 1981). Tansy only occured at the edge of this habitat.

Figure 6: A typical picture of the habitat types for A) rock grooves, B) short meadow, C) tall meadow and D) stony meadow. For a picture of a tansy in a rock crevice, see Figure 5 (dead tansy).

(6) **Exposition** – tansy genets were either shorter or higher than the surrounding vegetation. I assumed the exposition to have different influences on the presence of the

aphids. First, the probability of getting colonised by a winged morph was higher for exposed genets because of better visibility. Second, the wind had a stronger effect on exposed genets and third, hidden genets were surrounded by more insects and thus, by more possible predators. But not only the surrounding vegetation hid tansy genets, sometimes also surrounding rocks covered the genets. Additionally, I recorded if the genet was growing below a tree canopy. The different possibilities resulted in the following five categories: Exposed (1), exposed and under a tree canopy (2), covered by surrounding rocks (3), hidden in the vegetation (4) and hidden in the vegetation and under a tree canopy (5).

(7) **Orientation** – the different sides of an island vary in surface temperature, solar radiation and water availability. The north side has less solar radiation and is therefore colder and less sensitive to droughts than the south side (Häyrén, 1914). Especially the large islands lay mostly in an east-west direction. The direction of the winds was very variable in the summer half of the year but the most common and the most powerful was the wind from the south-west. In July and August warm winds from the east and south-east were frequent as well (Halkka, 1971). With the compass, I decided on which side of the island the tansy genet was growing; on the north, north-east, east, south-east, south, south-west, west, north-west or in the centre of the island.

(8) **Distance to the shore** – Genets growing close to the shore were more affected by the strong winds blowing from the sea; they had a higher risk of flooding and received spray water more often. This presumably prevented a possible colonisation or longer survival time of the aphids. On the other hand, I assumed that winged morphs that crossed the sea were more likely to colonise genets close to the shore. I roughly distinguished three categories: Genets growing less than 5 metres from the shore, genets growing at a medium distance to the shore (5 to 20 metres) and genets growing more than 20 metres away from the sea shore.

(9) **Distance to the next genet** – the distance to the next genet was measured in metres. I assumed that genets, which grew close to an occupied tansy, were more likely to get colonised by walking aphids than genets growing further away.

On the islands with more than 250 genets (Storlandet, Vindskären, Langskär and Flakaskär), except Halsholmen, the variables described above were collected for only a subset of genets. For the remaining genets, only the number of ramets were counted. The full set of variables was obtained for at least 174 genets, which were chosen randomly subject to the constraint that all plants with aphids and of each group at least one genet were included in the subsample.

3.4.2 Variables recorded for surveyed islands

Aphids occupied not all islands with tansy. To analyse which factors influence the presence of *M. fuscoviride* and *M. tanacetaria* on the islands, I collected the following variables for all 59 islands with tansy.

(1) **Island size** –a field vole study (Pokki, 1981) in the same area showed that extinctions were less common on large islands (Crone, 2001). In our system, the size of the islands did not correspond directly to the patch size available for the aphids. But the size of the islands was expected to have an influence on the resource availability, and thus indirectly on the patch size. Nevertheless, patch area and density can be positively or negatively correlated (Hanski, 1999). The sizes of the islands were taken from the Tvärminne island register[1].

(2) **Total number of genets** – theory predicts that the extinction risk of local populations decreases with increasing patch area because expected population size increases with patch area (Hanski, 1998). In this case, the resource availability is described by the number and density of ramets, genets and groups on an island.

(3) **Total number of ramets –** one genet consisted of one to many ramets. Aphids on one ramet were called a colony. The number of ramets was the actual number of available hosts for the aphids.

(4) **Total number of groups** – all tansy genets were grouped. All genets with less than five metres distance to another genet in the group were considered to belong to the same group. The distance of five metres was chosen as a distance an aphid can probably only cross by flight. The total number of groups was another measure for the resource availability but at the same time, it gave an idea about the spatial distribution of the genets on an island. A low number of groups on a large island indicates a clustered, whereas a high number indicates a more uniform distribution.

(5) **Distance to the next group** – for each group, the shortest distance in metres to the next group was measured. For the analysis, the mean distance to the next group was used. If there was only one tansy group on an island, the distance to the next group corresponded to the distance to the next island with tansy. If there were more than one group, then the mean of the distances was calculated.

(6) **Distance to the next group in clock-wise direction** – as a measure of the distribution of tansy on an island, the shortest distance to the next tansy group in the clock-wise direction was measured. For the analysis, the mean distance to the next group in clock-

[1] The Tvärminne island register is maintained by M. Lindström of Tvärminne Zoological Station

wise direction on an island was used. If only one group was found on an island, the distance to the next group in clockwise direction corresponded to the circumference of the island. If more groups were present, then the mean distance was calculated in metres.

(7) **Presence of trees** –trees were only present on islands with enough accumulated soil. This is mainly the case on older and thus presumably larger islands. In the succession of the vegetation, the first tree species present on an island is *Sorbus aucuparia,* which forms small populations. *Pinus silvestris* is the most common tree and often forms small forests. A few *Picea excelsa* are also present. If trees grow on an island, the velocity of soil accumulation increases, the soil conditions change and there is more protection from the wind and the radiation (Häyrén, 1914).

(8) **Number of trees** – the number of trees per island was estimated.

(9) **Presence of ants** – the presence of ants was noted during the visit. If no ants were detected moving on plants or rocks, small stones were turned to look for the presence of ants. The absence of ants on an island would probably prevent the establishment of *M. fuscoviride* (Stadler et al., 2001) because *M. fuscoviride* is dependent on ants (Flatt & Weisser, 2000).

From the above listed variables, I calculated the following parameters:

(10) **Genet density** – the number of genets on an island divided by the size of this island in hectars. These parameter gives a more accurate description of the resource availability per area.

(11) **Ramet density** – the number of ramets on an island divided by the size of this island in hectars.

(12) **Group density** – the number of groups on an island divided by the size of this island in hectars.

(13) **Number of different habitat types** – On each island, one to eight different habitat types were found (the habitat types are listed above). On the islands with more than 250 genets, this parameter was calculated from the subsample.

(14) **Simpson's diversity index of habitat types** – the number of different habitat types does not accord with the distribution of genets among the habitats. Therefore, two diversity indices were calculated. The weighted mean of the proportional habitat abundances was calculated with the Simpson's diversity index (Peet, 1974, Simpson, 1949). This index measured the probability that two genets selected at random from all genets on an island grew in the same habitat type.

eqn.1 $$D = 1 / \sum_{i=1}^{S} P_i^2$$

where D is the index, S the total number of different habitat types per island and P_i the proportion of total genets belonging to the *i*th habitat type. On the islands with more than 250 genets, this parameter was calculated from the subsample.

(15) **Shannon's diversity index of habitat types** – as in the Simpson diversity index, two components of diversity were combined, namely the number of different habitat types and the evenness of allotment of genets among the habitat types (Lloyd and Ghelardi, 1964).

eqn. 2 $$H = -\sum_{i=1}^{S} P_i \bullet \ln(P_i)$$

where H is the index, S the total number of different habitat types per island and P_i the proportion of total genets belonging to the *i*th habitat. 1/Exp (H) can be interpreted as the geometric mean of the proportional abundances (Hill, 1973). On the islands with more than 250 genets, this parameter was calculated from the subsample.

(16) **Most common habitat type** – the habitat type where most of the tansy genets per island were growing. On the islands with more than 250 genets, this parameter was calculated from the subsample.

(17) **Most common orientation** – the orientation on an island where most tansy genets were growing. On the islands with more than 250 genets, this parameter was calculated from the subsample.

(18) **Most common exposition** – the exposition where most tansy genets were found on an island. On the islands with more than 250 genets, this parameter was calculated from the subsample.

(19) **Most common phenology** – the phenological state which was the most frequent on an island. On the islands with more than 250 genets, this parameter was calculated from the subsample.

(20) **Mean phenological state** – the phenological states were ordered from shoot, bud, flower, withered, seed, crippled to dead. Then, the mean phenological state for each island was calculated. On the islands with more than 250 genets, this parameter was calculated from the subsample.

(21) **Most common distance to the shore** – the distance to the shore category where most genets were found. Thus, it revealed at which distance to the shore, the majority of genets on an island were growing. On the islands with more than 250 genets, this parameter was calculated from the subsample.

(22) **Number of genets growing less than five metres from the shore** – the number of genets which grew less than five metres from the shore, thus the number of genets belonging to distance-to-shore category 1 (see above). On the islands with more than 250 genets, this parameter was calculated from the subsample.

(23) **Mean height** – the mean of the maximum height in centimetres measured for each genet. Indicates the mean height of tansy on an island. On the islands with more than 250 genets, this parameter was calculated from the subsample.

(24) **Mean distance to the next genet** – the mean of the distance to the next genet in metres. On the islands with more than 250 genets, this parameter was calculated from the subsample.

(25) **Distance to the next island occupied by *M. fuscoviride* –** for each island, the distance to the next island occupied by *M. fuscoviride* in metres was measured. To obtain the distances, a map (Maastokartta, 2011 11 Tvärminne, Peruskartta 1:20 000, Maanmittauslaitos, 1999) was scanned, the centroids marked (Adobe Photoshop 5.5) and the distances between the centroids calculated (the programm has been written by M. Salathé). The centroids were chosen because they reflected the mean distance an aphid would travel in moving from one island to another (Crone et al, 2001).

(26) **Distance to the next island occupied by *M. tanacetaria*** – for each island, the distance to the next island occupied by *M. tanacetaria* in metres was measured. The procedure for obtaining the distances was the same as described for the distance to the next island occupied by *M. fuscoviride*.

(27) **Isolation index for *M. fuscoviride*** – metapopulation theory predicts that more isolated patches have a lower colonisation probability (Hanski, 1999). For each island an isolation index (I) was calculated as follows:

eqn 3
$$I = \sum_{j=1}^{j} P_j \bullet e^{(-\alpha d_{ij})} \bullet A_j$$

where $P_j = 1$ if the island $_j$ was occupied by *M. fuscoviride* and $P_j = 0$ if the island $_j$ had no aphids, d_{ij} is the distance between the island $_i$ and the island $_j$ in metres, A_j is the size of the island $_j$ in square metres (Hanski, 1994). The constant α is a species specific parameter of mobility. There are no estimation of α for aphids in the literature, therefore the α –value was estimated the way that the smallest isolation index had a value of 0.01. The α - value was set to 0.01. The distances between island $_i$ and islands $_j$ were calculated the way as it was described for the distance to the next occupied island.

(28) **Isolation index for *M. tanacetaria* –** the isolation index was calculated as the isolation index for *M. fuscoviride*. The only change was that $P_j = 1$ when the island $_j$ was occupied by *M. tanacetaria* instead of *M. fuscoviride*.

(29) **Distance to the mainland** – the distance from the centroid of an island to the closest point on the mainland was measured on a map (Maastokartta, 2011 11 Tvärminne, Peruskartta 1:20 000, Maanmittauslaitos, 1999) and converted into metres. In a mainland-island metapopulation, the distance to the mainland influences the colonisation of the islands. More distant islands are less likely to be colonised because migration takes place from the mainland to the islands and migrating individuals are assumed to have a constant probability of settling down which leads to a back-to-back exponential distribution (Neubert et al., 1995).

3.4.3 Analysis

All statistical analysis were performed with SPSS (Version 10.0, SPSS Inc. 1999). Averages are presented as mean ± standard error. If the P-value of an analysis is not mentioned explicitly, then the significant results were marked with one asterisk for $P<0.05$, with two asterisks for $P<0.01$ and with three asterisks for $P<0.001$ respectively.

Principal Component Analysis

Principal Component Analysis (PCA) were used to determine correlations between the variables. To select the number of extracted components, two criterions were used. The Kaiser criterion selects all components with an eigenvalue larger than 1. If less than 75% of the total variance was explained by the components selected with this criterion, the Joliffe criterion was applied. The Joliffe criterion selects all components with an eigenvalue larger than 0.7. Thus, it is less conservative than the Kaiser criterion (http://obelia.jde.aca.mmu.ac.uk).

The selected components were normalised with the Kaiser Normalisation and rotated with the Varimax method. The aim of the Varimax rotation is to remove, as far as possible, loadings in the mid range, e.g. in the range of 0.3 – 0.7 (http:// obelia.jde.aca.mmu.ac.uk). To detect multicollinearities, the component loadings were analysed. The component loadings are the correlation coefficients between the variables (rows) and the components (columns). The squared component loading is the percent of variance in that variable explained by the component. Component loadings were denoted high if larger than 0.6, medium if between 0.4 and 0.6 and low if smaller than 0.4. If different variables have similar component loadings, they are correlated. If a maximum of three components is selected, the easiest way to detect highly correlated variables is to plot the component 1 against the component 2. Variables, which lay closely together in this plot are highly correlated.

The nominal variables had to be converted into ordinal data because nominal data cannot be included in a PCA. The variables were ordered in the following way: **Exposition** – the ordinal scale followed the number of the different categories, starting from 1 for exposed to 5 for hidden in the vegetation and under a tree canopy. **Habitat type**- ordered from forest (1), rock crevice, rock groove, short meadow, dry stony meadow, stony meadow, wet tall meadow to tall meadow (8). This order was chosen because I assumed tansy quality to increase in this order. **Orientation**– most islands consisted mainly of a long south and north side, which differed as described in chapter 3.4.1. Thus, the orientations were associated to the following three categories: category 1: S, SW, SE (dry, sunny); category 2: E, W and centre; category 3: N, NW, NE (more humid, less radiation). **Phenology** – the phenological states were ordered from shoot (1), bud, flower, withered, seed, crippled to dead (7).

To stress some of the important correlations, the Spearman rank correlation coefficient (r_s) was calculated for a few pairs.

Logistic regression

The influence of the island variables on the presence of aphids on the islands was analysed with separate logistic regressions for *M. fuscoviride* and *M. tanacetaria*. The influence of the genet variables on the presence of *M. fuscoviride* and *M. tanacetaria* on genets was also analysed with separate logistic regressions for each species. For these analysis, only data from islands occupied by the respective species were used.

To keep the logistic models as simple as possible, a subset of the variables were chosen. Only variables which differed significantly between islands with and without aphids were selected for the logistic regression testing for aphid presence on the islands. For the logistic regressions testing for the presence of aphids on the genets, only variables which differed between genets with and without aphids were used. To evaluate differences, t-tests were performed for metric data with equal variances (Levene's test for equality of variance) and Mann-Whitney-U tests for metric data with unequal variances. Pearson chi-square tests were performed for nominal variables. For each set of selected variables, a PCA was performed to avoid multicollinearities between the variables entered in the logistic regression model. From each pair or group of correlated variables, the one with the lowest P-value from the t-test, Mann-Whitney-U or Pearson chi-square test was chosen for the logistic regression.

For the logistic regressions, the nominal and ordinal data had to be coded into dummy variables. The repeated method for the phenological states and the indicator method for the other categorical variables were used. The repeated method creates dummy variables in a way that each category except the first is compared to the category that precedes it. The order of the phenological states was the same as the one used for the PCA described above. The indicator method compares each category with a chosen reference category. For the indicator method the category with the most genets was chosen as the reference category, each other category was compared to this reference category (SPSS Inc., 1999).

To choose the best logistic model, two methods were used. First, the forward stepwise procedure was applied. The entry of a variable is tested using a 0.05 significance level of the score statistic. The necessity of removing a variable is tested based on the probability of a likelihood ratio statistic, based on the maximum partial likelihood estimates (SPSS Inc, 1999). Second, the Akaike information criteria (AIC, eqn. 4) was calculated.

eqn. 4 $$AIC=(-2 \cdot \text{log likelihood ratio}) + (2 \cdot \text{number of parameters})$$

The –2 log likelihood ratio is a sign for the fit of the model, the smaller this number the better the model fit. A model with more parameters has a higher chance to be a randomly and/or artificially improved model. Therefore the AIC adds to the –2 log likelihood the double of the number of parameters included in the model. The best model is the one with the lowest AIC. The AIC was calculated for a few models, which were likely to be good models. The comparison of the AIC gave an indication about the difference between these models.

For the interpretation of the logistic model, the odd ratios were analysed. The odds are defined as the probability that an event will occur divided by the probability that it will not occur (Motulsky, 1995). The odds ratio is the ratio of the odds, e.g. the odds of the colonised islands divided by the odds of the uncolonised islands. The odds ratio indicates the direction and the strength of a relationship. For a particular variable, an odds ratio of less than one corresponds to a decrease in the odds. An odds ratio of larger than one corresponds to an increase in the odds. Odd ratios close to one refer to a weak increase or decrease respectively in the odds.

Multiple linear regression

Multiple linear regressions were used to analyse the influence of the islands variables on the number and percentage of occupied ramets. These were performed separately for *M. fuscoviride* and *M. tanacetaria*. To keep the linear regression as simple as possible, a subset of variables were chosen. The procedure of choosing the subset of variables was the same as for the logistic regression.

To choose the best model, the forward stepwise procedure was used. This procedure enters a variable into the model when the corresponding P-value is lower than 0.05 and removes it from the model when the P-value exceeds 0.1 (SPSS Inc, 1999).