UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE BIOLOGIA ANIMAL

Genomic map of the past: Reconstructing the demographic history of wood ant species and their hybrids

Beatriz Cunha Portinha

Mestrado em Biologia Evolutiva e do Desenvolvimento

Dissertação orientada por:
Professor Doutor Vítor Sousa
Doutor Jonna Kulmuni

2020

Acknowledgements

My greatest gratitude must go to my parents. Without them, I would not be alive. Without their enrichment, encouragement, and support, I would not have had the opportunities I did or be where I am today. Thank you, mom and dad, for everything.

To my sister, Inês, thank you for existing and putting up with me when you feel like it. I appreciate you more than I can put into words. To Carlos, my brother-in-law, thank you for putting up with my sister, me, and the cat.

To my little Inês, thank you for choosing to grow with me and bringing me so much joy. It has been, and will continue to be, an honour to have you under my wing. To my friends, thank you for being wonderful companions and for sticking with me, even when I am annoying. My life is richer for having you in it.

I must thank my incredible supervisors. Thank you, Jonna, Vítor, and Pierre, for giving me the opportunity to develop such a cool project with an amazing system. I have learned so much from all of you. The knowledge and skills I gained throughout the course of this project, as well as the personal growth I was forced to undergo due to living alone for the first time in a country that was not my own, is something I will always take with me and cherish throughout my life and career.

To my SpecIAnt "labmates", Ina, Elisa, Jack, and Raphael, and to everyone in ESB, thank you for being so welcoming. You were truly wonderful. To Purabi, thank you for being my very first friend in Helsinki. You will always be dear to me. To Laura, Giovanna, Sara, and Camila, thank you for giving me a little taste of home when I needed it.

I would also like to thank the Erasmus+ programme for providing me with the opportunity and fundamental funding to develop my thesis at the Faculty of Biological and Environmental Sciences of the University of Helsinki, Finland. I would also like to thank *Societas pro Fauna et Flora Fennica* for the additional funding granted to this project.

Abstract

Hybridization is a process known to occur in various animal and plant *taxa*, leading to the combination of genetic material from previously isolated gene pools. While hybridization can lead to favourable outcomes, such as adaptive introgression, reconstructing speciation and hybridization histories is essential to better understand such outcomes. In Southern Finland, the distributions of the wood ant species *Formica polyctena* and *F. aquilonia* overlap and these species interbreed, producing viable and stable hybrid offspring. Whether these hybrid populations have a single origin or were formed through independent hybridization events remained an open question. In this project, we used genome-wide polymorphism data to study speciation and hybridization between *F. polyctena* and *F. aquilonia*. We characterized the genetic diversity and differentiation of populations of *F. polyctena* and *F. aquilonia* sampled across Europe, and hybrid populations from Finland. We modelled the demographic history of parental species and inferred their relationship with hybrid populations using the site frequency spectrum. To reconstruct the speciation history between *F. polyctena* and *F. aquilonia*, we considered alternative models of divergence, with and without gene flow. Across comparisons with different pairs of populations, we found that divergence between these species started in the Pleistocene, with continuous asymmetric gene flow from *F. aquilonia* into *F. polyctena*. The genomic patterns consistent with asymmetric migration could not be explained by gene flow from unsampled species, more closely related to *F. polyctena* or *F. aquilonia*. To reconstruct the hybridization history in Finland, we tested alternative secondary contact and admixture scenarios. Our results confirm that the three Finnish populations studied likely arose due to hybridization between *F. polyctena* and *F. aquilonia*. Our estimates indicate a higher contribution from *F. polyctena* into all hybrids (0.55 to 0.65 depending on the population), and strongly suggest that the two genetic lineages in Långholmen, the most extensively studied population of *F. polyctena* x *F. aquilonia* hybrids, share the same origin. It is, however, unclear whether this is the case for the remaining hybrid populations. This is the first study modelling the demographic history to elucidate the speciation of *rufa* group species. This allows us to provide insight into speciation with gene flow in eusocial haplodiploid organisms. Our findings concerning admixture between *F. polyctena* and *F. aquilonia* expand on the current knowledge on hybridization in the *rufa* group and will be useful to interpret the observed patterns of variation in *F. polyctena* x *F. aquilonia* hybrid genomes.

Keywords

Speciation; Hybridization; Demographic History; *Formica* ants

Resumo

A especiação é um processo que ocorre através da formação de mecanismos de isolamento reprodutivo que impedem a ocorrência de fluxo genético entre *taxa* divergentes. Quando o isolamento reprodutivo ainda não está completamente instalado, pode ocorrer hibridação caso ocorra contacto entre *taxa* divergentes. Caso os indivíduos híbridos sejam viáveis e se consigam reproduzir com as espécies parentais, isso pode levar à introgressão de alelos de uma espécie para outra. Isto é frequente quando populações de linhagens relacionadas que se encontravam isoladas voltam a estar em contacto, por exemplo, devido à remoção de barreiras físicas e/ou devido à expansão da área de distribuição de uma ou ambas as espécies. A hibridação é um processo que era considerado raro, mas que recentemente, com base em dados genéticos, se mostrou ser comum em vários *taxa* de plantas e animais. Ao nível genómico, a hibridação resulta na combinação de material genético proveniente de *gene pools* previamente isoladas. A partilha de alelos entre espécies pode desempenhar um papel na sua especiação, quer através da erosão das diferenças acumuladas entre elas, atrasando o processo da sua especiação, ou através da introdução de alelos com valor adaptativo em qualquer uma das espécies, o que pode acelerar a resposta adaptativa das populações, conhecido como introgressão adaptativa. Por outro lado, a combinação de alelos de linhagens distintas pode ter um efeito negativo na *fitness* dos híbridos devido a incompatibilidades genéticas. Devido à possibilidade de obter dados genómicos, tornou-se possível elucidar qual o papel da hibridação na especiação, nomeadamente para compreender a interação entre processos neutros (por exemplo, deriva e fluxo genético), incompatibilidades e genes envolvidos na adaptação. Enquanto que uma situação em que hibridação resulta em introgressão adaptativa é claramente benéfica para as espécies, é ainda difícil detectar com precisão os genes e regiões genómicas envolvidas. Um dos passos fundamentais para interpretar padrões genómicos é a reconstrução das histórias evolutivas de especiação e hibridação entre espécies.

No Sul da Finlândia, as distribuições de duas espécies próximas de formigas eusociais haplodiplóides, *Formica polyctena* e *F. aquilonia*, sobrepõem-se, pelo que estas espécies têm a possibilidade de se encontrarem e produzir descendência híbrida nesta região. De facto, estas espécies hibridizam nesta região e produzem populações estáveis de indivíduos híbridos viáveis, que são estudadas há mais de uma década. Apesar de estas populações terem sido geneticamente caracterizadas com o uso de vários marcadores moleculares, há várias questões que continuam em aberto: As diferentes populações híbridas têm uma origem comum ou resultam de múltiplos eventos independentes de hibridação? As espécies parentais divergiram há quanto tempo, e divergiram com ou sem fluxo genético?

Neste projeto, usámos dados genómicos obtidos através da sequenciação de genomas inteiros (*whole-genome sequencing*) para estudar a especiação e hibridação entre *F. polyctena* e *F. aquilonia*. Indivíduos de ambas as espécies parentais foram amostrados em diferentes pontos das suas distribuições na Europa (*F. aquilonia* na Escócia, *F. aquilonia* e *F. polyctena* na Suíça e Finlândia), e indivíduos híbridos foram amostrados em três locais no Sul da Finlândia (4 populações híbridas, devido à existência de duas linhagens genéticas distintas num dos locais de amostragem). A partir dos dados genómicos, foram genotipados um total de 59 fêmeas em aproximadamente 2.36 milhões de SNPs. Estes dados de polimorfismo foram utilizados para caracterizar a diversidade e diferenciação genética das populações. Um dos objectivos foi reconstruir a história demográfica das populações destas espécies comparando diferentes modelos para testar hipóteses sobre a divergência e hibridação, utilizando dados com base no espectro de frequências alélicas (*site frequency spectrum*).

Os resultados indicam que tanto populações das espécies parentais como de populações híbridas possuem relativamente pouca diversidade genética, visto que as estimativas de heterozigotia esperada são menores que 0.19 em todos os casos. As espécies parentais estão bastante diferenciadas uma da

outra, com valores de $F_{ST} > 0.25$ em todos os casos. As populações híbridas aparentam ser geneticamente intermédias entre as suas espécies parentais, apresentando menor diferenciação genética com a população de *F. polyctena* da Finlândia. No entanto, indivíduos híbridos são mais semelhantes entre si do que a indivíduos de qualquer uma das espécies parentais. Todas as populações híbridas apresentam valores de F_{IS} negativos entre -0.08 e -0.245, indicando desvios ao equilíbrio de Hardy-Weinberg consistentes com cruzamentos entre linhagens distintas (*outcrossing*) recente.

Para estudar a especiação entre *F. polyctena* e *F. aquilonia*, testámos modelos demográficos alternativos que consideraram cenários com e sem fluxo genético entre as espécies. Com base no SFS observado, foi possível calcular a verossimilhança de cada modelo, assim como estimar os respectivos parâmetros (por exemplo, efectivos populacionais, tempo de divergência, taxas de migração). Estes modelos foram testados com diferentes pares de populações em que as populações podem estar geograficamente distantes, como no caso da comparação entre a população de *F. polyctena* amostrada no Oeste da Suíça e a de *F. aquilonia* amostrada na Escócia, ou próximas, como na comparação entre populações de ambas as espécies amostradas na Suíça. Transversalmente às diferentes comparações, as nossas estimativas sugerem que a divergência entre *F. polyctena* e *F. aquilonia* começou no Pleistoceno (entre 517,580 e 743,078 anos atrás, dependendo do par de populações considerado) e que ocorreu com fluxo genético assimétrico contínuo de *F. aquilonia* para *F. polyctena* (com cerca de 0.57 a 1.4 migrantes por geração a migrarem de *F. aquilonia* para *F. polyctena*, dependendo do par de populações considerado). De modo a verificar que este aparente fluxo genético assimétrico não é devido a introgressão com outras espécies, considerámos modelos com populações não amostradas. Os resultados indicam que os padrões genómicos não podem ser explicados por fluxo genético entre *F. polyctena* ou *F. aquilonia* com outras espécies não amostradas e geneticamente mais próximas, e que modelos com fluxo genético diretamente de *F. aquilonia* para *F. polyctena* explicam melhor os dados de SFS observados. Quando testámos os mesmos modelos com as populações de *F. polyctena* e *F. aquilonia* amostradas na Finlândia, obtivemos parâmetros muito semelhantes, sugerindo a mesma história evolutiva no geral. No entanto, também encontrámos uma diferença importante, dado que nas populações da Finlândia as estimativas suportam fluxo genético bidirecional, com migração a ocorrer maioritariamente de *F. aquilonia* para *F. polyctena* (1.28 migrantes por geração, o que é semelhante ao obtido com os restantes pares de populações), mas também com algum fluxo genético de *F. polyctena* para *F. aquilonia* (0.2 migrantes por geração). Este fluxo poderá acontecer de forma direta através do aumento das oportunidades de contacto nesta região devido à alteração artificial dos habitats destas espécies, ou de forma indirecta via fluxo genético entre híbridos e indivíduos de *F. aquilonia*.

Para estudar a origem de cada população híbrida que foi amostrada no Sul da Finlândia, comparámos cenários de contacto secundário, em que a população considerada "híbrida" teria na verdade divergido mais recentemente de uma ou ambas as populações parentais, seguido por contacto secundário com a população ou populações parentais mais distantes. Estes foram comparados com cenários em que a população híbrida é originada por hibridação entre *F. polyctena* e *F. aquilonia*. Em todos os casos, os nossos resultados confirmam que hibridação entre estas espécies é a explicação mais provável para a origem das populações híbridas, dado que modelos com contacto secundário obtiveram valores de verossimilhança estatística mais baixos. As estimativas dos nossos modelos indicam que a contribuição genética de *F. polyctena* para as populações híbridas é superior à de *F. aquilonia*, variando entre 0.55 e 0.65, dependendo da população híbrida considerada. Ainda, os nossos resultados sugerem que as duas linhagens genéticas distintas que existem em Långholmen, a população híbrida estudada mais extensivamente, foram muito provavelmente originadas pelo mesmo evento de hibridação entre *F. polyctena* e *F. aquilonia*, partilhando várias gerações de ancestralidade comum. Apesar de os nossos resultados indicarem que este cenário de uma origem híbrida única também é o mais provável para as restantes populações híbridas amostradas, as estimativas desse modelo não apoiam essa hipótese, dado

que sugerem que o tempo de divergência coincide praticamente com o tempo de hibridação. Estas estimativas são, por isso, compatíveis com origens múltiplas independentes que tenham ocorrido na mesma época e com contribuições semelhantes de ambas as espécies parentais.

Este projeto é o primeiro em que modelação demográfica foi utilizada para estudar a especiação entre espécies do grupo *rufa* do género *Formica*. Assim, as nossas conclusões quanto à história de especiação entre *F. polyctena* e *F. aquilonia* permitem avançar a compreensão de especiação com fluxo genético assimétrico entre espécies de organismos haplodiplóides eusociais. Os nossos resultados relativos à hibridação entre estas espécies e à sua descendência híbrida ampliam o conhecimento já existente referente à hibridação no grupo *rufa*, e serão úteis na interpretação dos padrões observados de variação genética em genomas híbridos entre *F. polyctena* e *F. aquilonia*.

Palavras-chave:

Especiação; Hibridação; História Demográfica; Formigas *Formica*

Index

Index of Tables and Figures

Abbreviations

bp – base pair

VCF – Variant Call Format

HWE – Hardy-Weinberg Equilibrium

SFS – Site Frequency Spectrum

LD – Linkage Disequilibrium

H_e – Expected Heterozygosity

H_o – Observed Heterozygosity

F_{IS} – Inbreeding Coefficient

F_{ST} – Fixation Index

PCA – Principal Component Analysis

PC – Principal Component

K – Number of ancestral clusters

1. Introduction

Speciation is the process that leads to the establishment of reproductive isolating mechanisms that prevent gene flow between newly emergent taxa (Coyne and Orr, 2004). This process can take place in different modes often defined by the spatial context (Mallet *et al.*, 2009): sympatric (in which individuals are physically capable of meeting with fairly high frequency), parapatric (in which populations occupy distinct but contiguous geographic regions; only a small fraction of individuals from different populations will meet), or allopatric (in which populations are separated by uninhabited space across which dispersal is very limited or non-existent). Until reproductive isolation has completely developed between diverging populations, hybridization and introgression are still possible in geographical contexts that allow individuals to meet. In fact, hybridization is known to occur in nature across many animal and plant *taxa*. This process leads to the combination of genetic material from divergently-adapted gene pools by the interbreeding of genetically distinct populations (Schwenk, Brede and Streit, 2008). Accordingly, hybridization can take place between populations of the same species (i.e., intraspecific hybridization) or between populations of different species (i.e., interspecific hybridization). In addition to being recognized as a driver of speciation, in the so-called hybrid speciation (where new hybrid populations become isolated from their parental populations and give rise to a new species; Baack and Rieseberg 2007) and in instantaneous speciation (Mallet, 2007), hybridization has also been proposed as an important component in different modes of non-allopatric speciation (Abbott *et al.*, 2013) and in the reinforcement of species barriers (Mallet, 2007).

Hybridization between species can allow for the introgression of alleles from one species into the other, therefore playing a role in speciation. This can happen either by eroding the divergence between species and therefore slowing the speciation process, or by introducing useful alleles for the colonization of novel habitats, thus enabling local adaptation that can lead to divergence, and eventually speciation. However, introgression is expected to vary along the genome, with limited introgression in genomic regions with incompatible loci. Furthermore, modern genomic patterns of admixture may be a single snapshot of complex interactions among divergent populations that continuously change through time and space (Abbott *et al.*, 2013). The demographic history of populations (such as their times of divergence, changes in effective size, and levels of gene flow between populations) leaves signatures in genome-wide polymorphism patterns. Thus, we can use genomes to reconstruct key past demographic events, including quantifying historical levels of gene flow (Sousa and Hey, 2013; Beichman, Huerta-Sanchez and Lohmueller, 2018). This allows us to understand how demography shapes genomic divergence during speciation (Welch and Jiggins, 2014). Moreover, the effects and efficiency of natural selection are heavily affected by the demographic history of populations, particularly by past effective population sizes, migration rates and times of split (Sousa and Hey, 2013). Thus, knowing the demographic history of hybrid populations is instrumental to understand hybridization and patterns of introgression, as well as to detect regions under selection, either because they are involved in adaptation or due to incompatibilities.

In Southern Finland, the distributions of two wood ant species of the *Formica* genus, *Formica polyctena* and *F. aquilonia*, overlap and these species coexist. Both species are known to have vastly polygynous nests, i.e., nests with hundreds of queens (Pamilo, 1982). These haplodiploid arrhenotokous ants (i.e., males are haploid and females are diploid; mothers monopolize the production of male offspring by the asexual production of sons; De La Filia, Bain and Ross, 2015) can be classified as habitat specialists and both form large polydomous societies (i.e., associations of cooperating nests) in coniferous and mixed forests (Pamilo, 1982). Hybridization between these species has led to the formation of several hybrid populations, of which the most studied is located in Långholmen, Hanko Peninsula. This hybrid population comprises individuals that are morphologically intermediate between *F. polyctena* and *F.*

aquilonia, and was found to contain two distinct hybrid lineages with large-scale intersexual genetic differences (W lineage is widespread in the population, and R is rare), with males forming two highly divergent gene pools (Kulmuni, Seifert and Pamilo, 2010). Using amplified fragment length polymorphism (AFLP), Single Nucleotide Polymorphism (SNP), microsatellite and allozyme markers, the two gene pools were found to present broad genetic differences, indicating that recent gene flow between them is limited. Several studies have found signatures of contemporary selection in the hybrids, which may depend on temperature (Martin-Roy, Nygård *et al.*, submitted) and differences in ploidy levels between sexes (Kulmuni and Pamilo, 2014).

Beresford *et al.* (2017) further investigated hybridization between *F. aquilonia* and *F. polyctena* across 16 localities in Southern Finland, sampling more than 600 workers over a period of around nine years. While previously only a hybrid population with wholly *F. polyctena*-like mitochondrial haplotypes had been documented (Kulmuni, Seifert and Pamilo, 2010), this study identified new populations with exclusively *F. aquilonia* mitochondrial haplotypes, as well as locations where both *F. polyctena* and *F. aquilonia* mitochondrial haplotypes were present. A pattern of cytonuclear mismatch was also identified in the hybrids, in which nests with nuclear genomes closer to parental-like *F. polyctena* are more likely to have *F. aquilonia* mitochondrial haplotypes, and vice-versa. Incompatibilities between the nuclear and the organellar genomes may arise from cytonuclear mismatches in hybrid individuals (Burton, Pereira and Barreto, 2013), however, in this system, the possible incompatibilities are likely to be weak as they are not erased from the populations. This pattern of cytonuclear mismatch is unlikely to happen in a scenario of random mating without selection. The authors propose two hypotheses as to why this is observed: (1) females hold a preference for mating with heterospecific males or (2) they mate randomly but progeny with heterospecific cytonuclear combinations are favoured (i.e., have higher fitness) than those with conspecific combinations. Over half of the localities presented signatures of admixture and different localities were found to exhibit patterns of genetic variation consistent with several hybridization events or, alternatively, with having backcrossed with the parental species to different extents. Together with the fact that different hybrid populations across Southern Finland possess different mitochondrial haplotypes, which in itself suggests multiple admixture events, it is likely that hybridization between these species has been ongoing in this area for many generations.

Considering the well-characterized Långholmen population, Ghenu *et al.* (2018) developed a two-locus mathematical model with hybrid incompatibility, female heterozygote advantage, recombination (different levels of recombination were considered, in a range of 0 to 0.5) and assortative mating, emulating a scenario where hybridization is simultaneously favoured and selected against. This is what is observed in the hybrid populations, where it is advantageous for the females to be hybrids, but detrimental to the males (Kulmuni and Pamilo, 2014). This two-locus model resulted in a rugged fitness landscape where heterozygote genotypes have a higher fitness and incompatible double homozygotes do not experience this increased fitness. In agreement with what was found in the natural population (Kulmuni, Seifert and Pamilo, 2010; Kulmuni and Pamilo, 2014), the model predicts that males have reduced fitness and survive better if one parental haplotype is fixed. Females suffer from the same incompatibility, but this is masked since they are diploid (in the case of a recessive incompatibility). Therefore, diploid females take their maximum profit from heterozygosity. While this model is relatively simple and more complex models with more than two incompatible loci would be better able to mimic the natural hybrid populations, the authors predict that the Långholmen population may be moving towards a scenario that is mediated by high frequencies of introgressed females. Hence, it may be approaching a favourable outcome in which there is a compromise between male and female interests. However, in order to understand the maintenance of polymorphisms and the dynamics of the compromise between males and females, we need information on the demographic history of the Långholmen population, particularly about key events such as the levels of past gene flow and the time

of divergence from the parental species. Recent methods (e.g., Excoffier *et al.*, 2013) allow us to date the divergence of populations and quantify past levels of gene flow from the site frequency spectrum (SFS), which can be obtained from genome-wide data and describe the distribution of allele frequencies in a sample. This method also allows comparing the fit of the data to alternative models, which can represent alternative modes of divergence, e.g. divergence without gene flow followed by secondary contact *versus* divergence with gene flow.

Computing the SFS is an effective approach towards summarizing the within- and between-populations variation contained in genome-wide data. The SFS can be computed with information of only one population (1D-SFS) or using data from two or more populations (multidimensional SFS). Excoffier *et al.* (2013) developed a coalescent SFS-based composite-likelihood method to infer the past demography of a set of populations from large genomic datasets, implemented in the fastsimcoal2 software. By approximating the expected SFS from simulations under complex demographic models, fastsimcoal2 can find the set of parameter estimates that maximize the likelihood of a given model. In recent studies, this software has been successfully used to study divergence between species (e.g., Oswald *et al.*, 2017; Hotaling *et al.*, 2018), as well as hybridization (e.g., Filatov, Osborne and Papadopulos, 2016; Chan *et al.*, 2017; Ru *et al.*, 2018). In essence, fastsimcoal2 is sufficiently powerful to disentangle the effects of similar scenarios where gene flow takes place, such as speciation with gene flow *versus* secondary contact (Filatov, Osborne and Papadopulos, 2016) and to reliably reconstruct complex evolutionary histories of related species (Oswald *et al.*, 2017).

In this work, we used genome-wide genomic data to study the speciation history between *F. polyctena* and *F. aquilonia* both 1) outside and 2) within Finland. The goal is to answer the following questions: Did the species diverge in allopatry or with gene flow? Is there evidence of population size changes in either species? What is the timing and number of demographic events? What are the estimated population sizes and timing of divergence? Is there greater support for gene flow between the species in Finland (where distributions overlap) compared to central Europe? Is the history of population size changes more complex in Finland? Due to genetic and morphological evidence that other species within the group can hybridize with both our study species (e.g., Seifert, Kulmuni and Pamilo, 2010), we also investigated if 3) there is evidence for gene flow from an unsampled species to either *F. aquilonia* or *F. polyctena*. To answer these questions, we tested 2-populations models with one population of each parental species, as well as 2-populations models with an additional unsampled population that represented, in turn, a sister species of each of the species under study in the present work (Goropashnaya *et al.*, 2012). Having knowledge on relevant parameters such as times of divergence, current and ancestral effective sizes, and levels of gene flow between the populations will not only elucidate on the history between them, but will also provide valuable insight to interpret summary statistics and population structure analyses that allow us to characterize present-day populations.

Lastly, we 4) studied the origin of the hybrid populations in Finland. Did they arise from admixture between *F. aquilonia* and *F. polyctena*? When was each hybrid population formed (i.e., what is the timing of each admixture event leading to the formation of each hybrid population)? Ultimately, did the different hybrid populations arise from independent hybridization events or from a single event (followed by subsequent divergence events and colonization of new geographical locations)? For this, we tested 3- and 4- population models where we studied each hybrid population alone with its putative parental populations (3-population models), or where we considered pairs of hybrid populations together with their parental populations (4-population models).

This project is the first to employ whole-genome polymorphism data to study speciation and hybridization in *Formica* ants. We found that *F. polyctena* and *F. aquilonia* diverged with asymmetric

gene flow, with migration occurring predominantly from *F. aquilonia* into *F. polyctena*. Our results strongly support the long-standing hypothesis that the hybrid populations observed in Southern Finland result from admixture between *F. polyctena* and *F. aquilonia* and propose that the W and R lineages in Långholmen share the same origin.

2. Material and Methods

2.1. Sampling

Sampling was carried out by Jonna Kulmuni, Pierre Nouhaud and collaborators prior to the beginning of my thesis. Females of each parental species were sampled from several locations across Europe (Fig. 2.1). For *Formica polyctena*, sampling was carried out in two locations in Switzerland and in the Åland islands, Finland, where three individuals were collected in each site. For *F. aquilonia*, individuals were collected from Switzerland, Scotland, and Central Finland. Sample sizes for these populations are also three individuals. In addition, one more female of each species was collected from locations in close proximity to the hybrid populations in Finland. Overall, 10 females were sampled for each species.

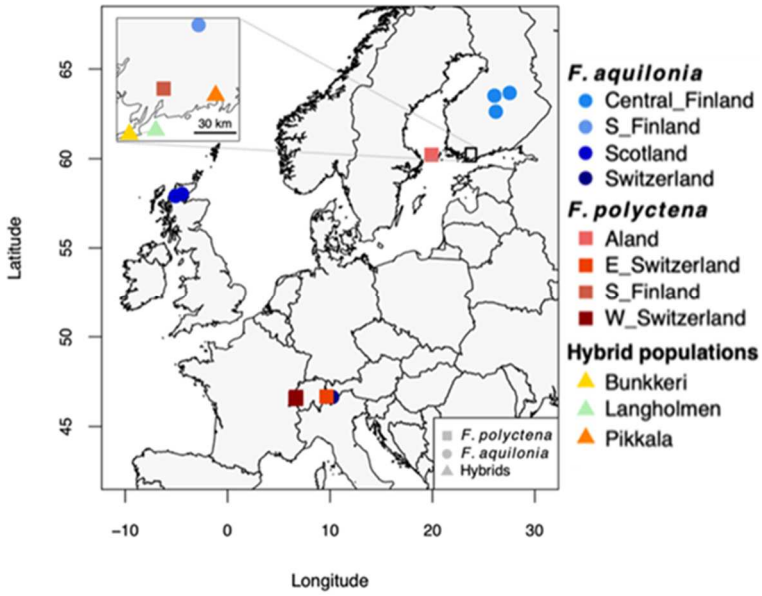


Figure 2.1 - Map of sampling locations. Each symbol represents a sampled individual (some are overlapping). Circles are used to represent *Formica aquilonia* individuals and squares represent *Formica polyctena*, while triangles are used for hybrid individuals.

Individuals from three known hybrid locations were sampled in Southern Finland (Fig. 2.1). Ten hybrid females were sampled from the Pikkala and Bunkkeri populations, while 10 females were sampled from each of the Långholmen lineages (R and W). For the purpose of our analyses, we considered the two lineages in Långholmen as two separate populations. Overall, 40 hybrid females were sampled.

2.2. DNA extraction and sequencing

DNA extraction and sequencing were performed by Novogene prior to the start of my thesis. DNA was extracted from whole-bodies with a SDS (sodium dodecyl sulfate) protocol. DNA libraries were constructed using NEBNext DNA Library Prep Kits (New England Biolabs). Samples were processed and sequenced at Novogene (Hong Kong) as part of the Global Ant Genomics Alliance (Boomsma *et al.*, 2017) which aims to sequence several hundred ant genomes. Whole-genome sequencing was performed on Illumina Novaseq 6000 (150 base pairs paired-end reads; aiming for 10x average coverage for diploid females and 5x for haploid males). In paired-end sequencing, both ends of a DNA fragment are sequenced. This method enables more accurate read alignment and increases indel (a type of variation where a nucleotide sequence is either present, through an insertion, or absent, through a deletion; Rodriguez-Murillo & Salem, 2013) detection (Goodwin, McPherson and McCombie, 2016).

Raw Illumina reads and adapter sequences were trimmed using Trimmomatic (v0.38; parameters LEADING:3, TRAILING:3, MINLEN:36; Bolger, Lohse and Usadel, 2014) before mapping against the reference genome (272 Mbp, 27 pseudo-chromosomes, Nouhaud *et al.*, in prep.) using BWA MEM with default parameters (v0.7.17; Li and Durbin, 2010). Duplicates were removed using Picard Tools with default parameters (v2.21.4; http://broadinstitute.github.io/picard).

2.3. SNP calling and filtering

Single nucleotide polymorphisms (SNPs) and genotypes were called with freebayes (v1.3.1; Garrison and Marth, 2012), disabling population priors (-k).

After calling the SNPs, we obtained a Variant Call Format (VCF) file which underwent comprehensive filtering. Various steps were taken to establish a high-confidence set of variants, the first of which was to remove sites that are at a distance of less than two base pairs (bp) from indels. Furthermore, we also removed SNPs that were only supported by Forward or Reverse reads. Only biallelic SNPs with quality equal or higher than 30 were kept. In order to refrain from removing entire sites when only a subset of individuals had inadequate genotype calls, individual genotypes with genotype qualities lower than 30 were coded as missing data. Genotypes with depth of coverage lower than eight were also coded as missing data. In addition, sites with missing data across more than 50% of the 100 individuals in our sample were removed. To avoid biases due to different forms of natural selection, for the demographic history analysis, we removed the third chromosome, also known as the social chromosome. This chromosome harbours genes responsible for polymorphism in social organization in *Formica* species, controlling if a colony is headed by one (monogynous) or multiple (polygynous) queens (Brelsford *et al.*, 2020). Recombination is rare between monogynous and polygynous versions of this chromosome (supergene, Brelsford *et al.*, 2020), leading to the maintenance of ancestral polymorphisms across *Formica* species which could bias our demographic inference.

To remove mapping errors that cause sites to show excessive heterozygosity (e.g., sites that are duplicated in all or some individuals in our sample but not in the reference genome or show excess coverage due to poorly mapped reads in repetitive regions), we applied a filter based on Hardy-Weinberg Equilibrium (HWE). We pooled all individuals together, regardless of their population of origin, purposefully creating excessive homozygosity via Wahlund effect (i.e., the apparent excess of homozygotes and the deficit of heterozygotes observed due to the existence of population subdivision; Garnier-Géré and Chikhi, 2013). This made it possible to identify and remove sites that were still excessively heterozygotic across all sampled individuals.

Lastly, we applied a filter based on individual coverage which enabled us to remove low-confidence sites due to low coverage, as well as potentially duplicated sites that have very high coverage. Instead of applying the same minimum and maximum coverage thresholds to every individual, we set individual-specific thresholds based on mean coverage values. While the lower bound of this interval is half the mean coverage of the individual in question, the upper bound corresponds to twice the mean coverage. For each individual, only sites whose coverage fell inside their interval were kept.

After preliminary analyses, one hybrid female was not assigned to the Långholmen R lineage from which it was collected. Since it could be a recent migrant that would bias results, we removed this individual from our population structure and demography analyses. At the end of filtering, we were left with 2 362 358 sites across all 59 remaining individuals.

2.4. Data analysis

2.4.1. Population structure

The genomic data was used to characterize the populations by computing summary statistics and analysing population structure. This was done in order to confirm the assignment of individuals into populations and to guide our interpretation of the demographic modelling results. All analyses pertaining to this matter were carried out with the R Software for Statistical Computing (v3.6.3; R Core Team, 2017).

Population structure was studied by means of two individual-based methods, Principal Component Analysis (PCA) and sNMF analysis (Frichot *et al.*, 2014), the latter of which estimates ancestry coefficients of individuals. These analyses were performed with custom-made scripts that employ the SNPRelate (v1.20.1 ; Zheng *et al.*, 2012) and LEA packages (v3.0.0; Frichot and François, 2015).

Summary statistics, such as observed and expected heterozygosity, inbreeding coefficients (F_{IS}) and pairwise fixation indexes (F_{ST}; computed using the Weir & Cockerham estimator; Weir and Cockerham, 1984) were calculated using custom-made scripts provided by Dr. Vítor Sousa and adapted by the present author. Pairwise F_{ST} values were also computed using the SNPRelate package.

2.4.2. Demographic modelling

Several alternative demographic models were designed and compared to answer our questions, which we tested using the site frequency spectrum (SFS) from different combinations of populations. We used fastsimcoal2 (Excoffier *et al.*, 2013), a composite-likelihood method, to test alternative models and infer demographic parameters (detailed in Supplementary Table 1) from the SFS. Each model was run 100 times, with 80 iterations for likelihood maximization and 200,000 coalescent simulations to approximate the expected SFS. The mutation rate was assumed as 3.5×10^{-9} (an average of mutation rates of various species from the Hymenoptera order; Liu *et al.*, 2017).

In *Formica*, young queens start laying eggs in their first years of life, however these eggs are likely to be reared into workers. The average age at which the queens start producing sexuals, and therefore contributing to the next generation, is two to three years. As such, generation time was assumed to be 2.5 years. After obtaining point parameter estimates and expected likelihoods for all models (detailed in section 2.4.2.1), tested with all population comparisons considered (detailed in section 2.4.2.2), the model with the highest expected likelihood was chosen as the best model in each case.

2.4.2.1. Model characteristics

2.4.2.1.1. Models to study the speciation history between *F. polyctena* and *F. aquilonia*

To answer our first two questions, "What is the speciation history between *F. aquilonia* and *F. polyctena*?" and "Is the history of divergence between *F. aquilonia* and *F. polyctena* different in Finland compared to Europe?", we created four 2-population models (Fig. 2.2) representing plausible modes of speciation between these two species. To serve their purpose, these models compared one population of *F. polyctena* to another of *F. aquilonia* (population comparisons are detailed in section 2.4.2.2).

Our first 2-population model is the "Allopatry" model (Fig. 2.2A), which considers that the populations remain isolated since their divergence until present time. This model also allows for the populations to change size, either expanding or contracting, happening at a time between the divergence of the populations and the present. Conversely, the "Sympatry" model (Fig. 2.2B) considers that the

populations are in constant contact since their divergence until present time, meanwhile also considering the possibility of a population resize. The “Isolation after Migration" model (Fig. 2.2C) allows the populations to exchange migrants after their divergence before a complete barrier (either physical or reproductive) to gene flow later becomes established, isolating the populations until present. In the “Migration after Isolation " model (Fig.2.2D), there is a period of isolation after initial population divergence, followed by removal of the barrier that isolated them, allowing the populations to exchange genetic material until present. In the “Isolation after Migration” and “Migration after Isolation” models, the populations may experience resizes at the time when the barrier to gene flow is removed.

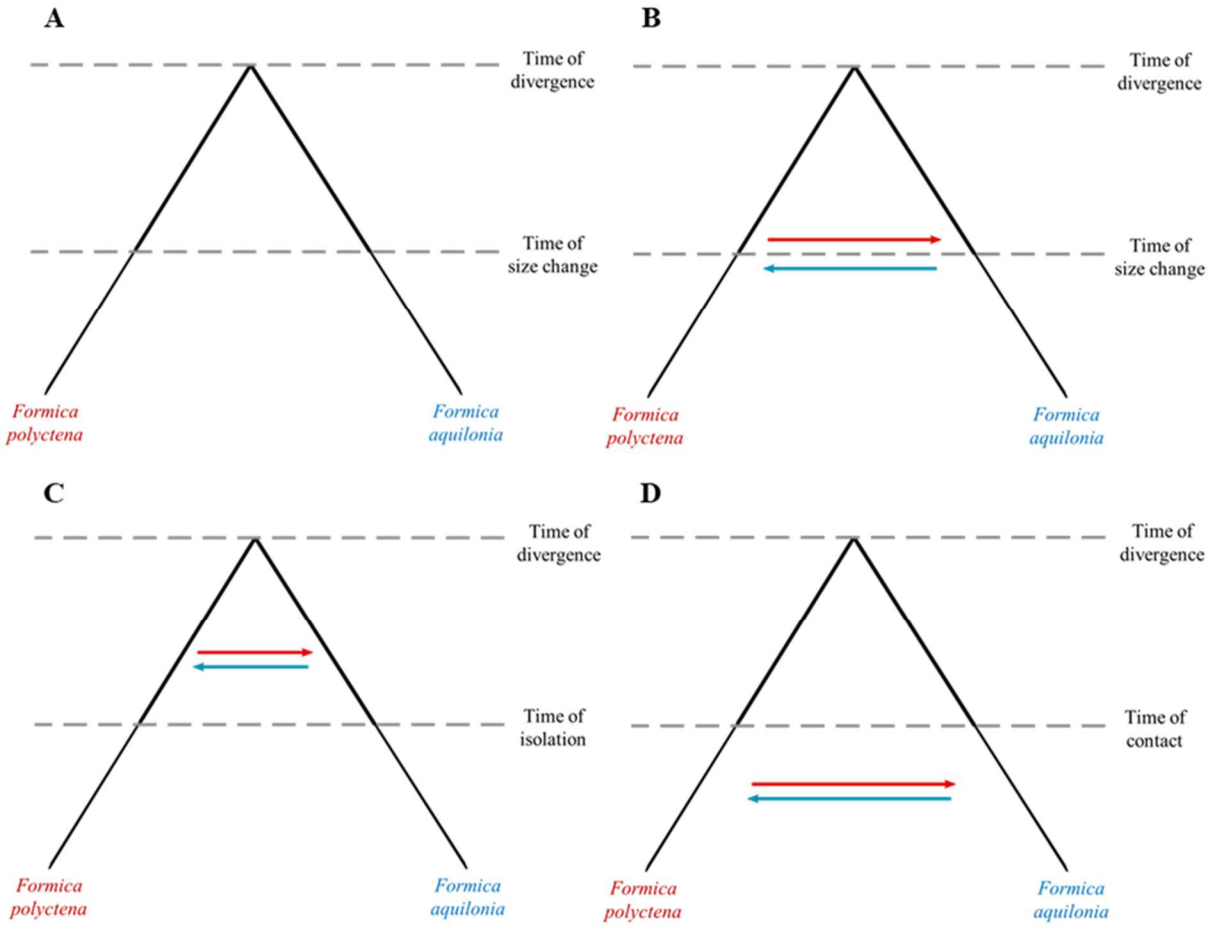


Figure 2.2 – Demographic models designed to study the speciation history between *Formica polyctena* and *Formica aquilonia*. **A** “Allopatry”: the populations diverge without contact. **B** “Sympatry”: the populations diverge with gene flow. **C** “Isolation after Migration”: after divergence, the populations exchange migrants until a barrier to gene flow is established. **D** “Migration after Isolation”: after divergence without contact, the barrier to gene flow disappears and the populations are free to exchange genetic material. Arrows represent migration. The direction of gene flow is indicated by the direction and color of the arrows (red represents gene flow from *F. polyctena* into *F. aquilonia*; blue represents gene flow from *F. aquilonia* into *F. polyctena*). The different thickness in the lines representing the populations represent changes in effective size, which can happen either by contractions or expansions.

2.4.2.1.2. Models to test for gene flow from an unsampled, closely related species

For our third question, “Is there evidence for gene flow from unsampled species to either *F. aquilonia* or *F. polyctena*?”, we introduced a “ghost” (i.e., unsampled) population into a model similar to the “Allopatry” model (Fig. 2.3). Based on previous knowledge, we assumed that if any gene flow between one or both of our sampled species and a third, unsampled species, were to exist, the unsampled donor species would most likely be a sister species of the receiver species (which could be any of our sampled species).

After diverging, the unsampled species would then send migrants to its sister species. Based on the phylogeny presented in (Goropashnaya *et al.*, 2012), and under such scenario, these unsampled species would be *F. rufa*, sister to *F. polyctena*, or *F. lugrubris* and/or *F. pratensis*, both more closely related to *F. aquilonia.*

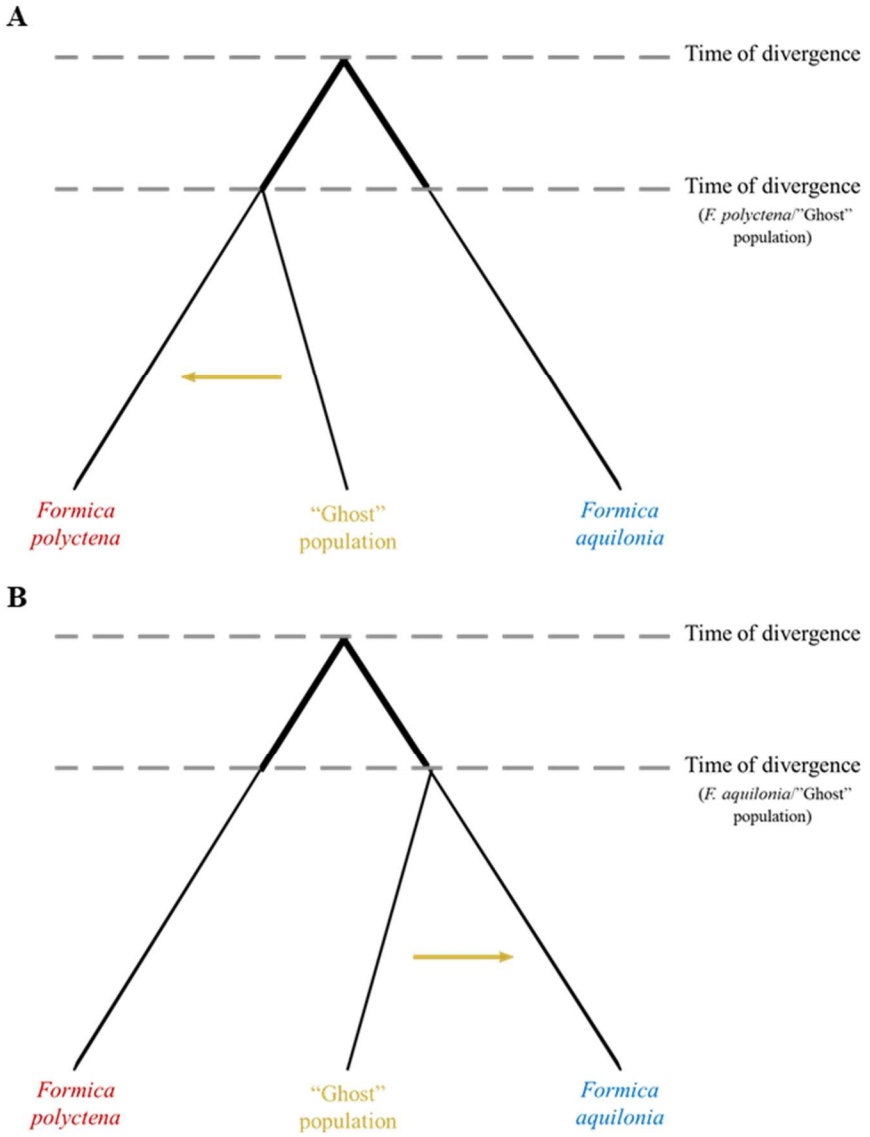


Figure 2.3 – Models designed to study possible introgression from an unsampled species ("ghost") into *Formica polyctena* or *Formica aquilonia*. **A** "((*F. polyctena*, Ghost population), *F. aquilonia*)": the "ghost" population represents a sister species of *F. polycena,* into which it sends migrants ever since their split until present. **B** "((*F. aquilonia*, Ghost population), *F. polyctena*)": the "ghost" population portrays a sister species of *F. aquilonia,* into which it sends migrants from the time of their split until present time. Gene flow and changes in population size are depicted as in Figure 2.2.

Accordingly, we built two models to study this matter. The "(*F. polyctena*, Ghost population), *F. aquilonia*" model (Fig. 2.3A), considers that the *F. aquilonia* population first diverges from the ancestral population of *F. polyctena* and the "ghost" population, followed by the divergence between these two populations. From this time until present, the "ghost" population will send migrants into the *F. polyctena* population. "(*F. aquilonia*, Ghost Population), *F. polyctena*" (Fig. 2.3B) instead considers that the *F. polyctena* population is the first to diverge, followed by the divergence between the *F. aquilonia* and the "ghost" population, after which the "ghost" population will send migrants into the *F. aquilonia* population until present.

2.4.2.1.3. Models to study the origins of the hybrid populations

For our final question, "How did the hybrid populations originate?", we designed 3-population and 4-population models. In the 3-population models, two are the Finnish parental populations, one from each species, and the third is a hybrid population. We assumed the single individuals collected in close proximity to the hybrid populations in Southern Finland to be the most adequate representatives of the parental species out of those sampled in this area. Therefore, these solitary samples were used as the parental populations of the hybrid populations.

These models were tested for each sampled hybrid population. As secondary contact scenarios can produce the same patterns of variation and differentiation as hybrid origin scenarios, we tested models where the putative hybrid population has, in turn, diverged from each of the parentals, as well as a trifurcation model where the parental and the hybrid populations diverge simultaneously (Fig. 2.4). All these scenarios include subsequent secondary contact after a period of post-divergence isolation. The hybrid population engages in secondary contact either with both parental populations, if the split between all populations was simultaneous, or with the more distant parental population, when the three populations do not split from each other at the same time.

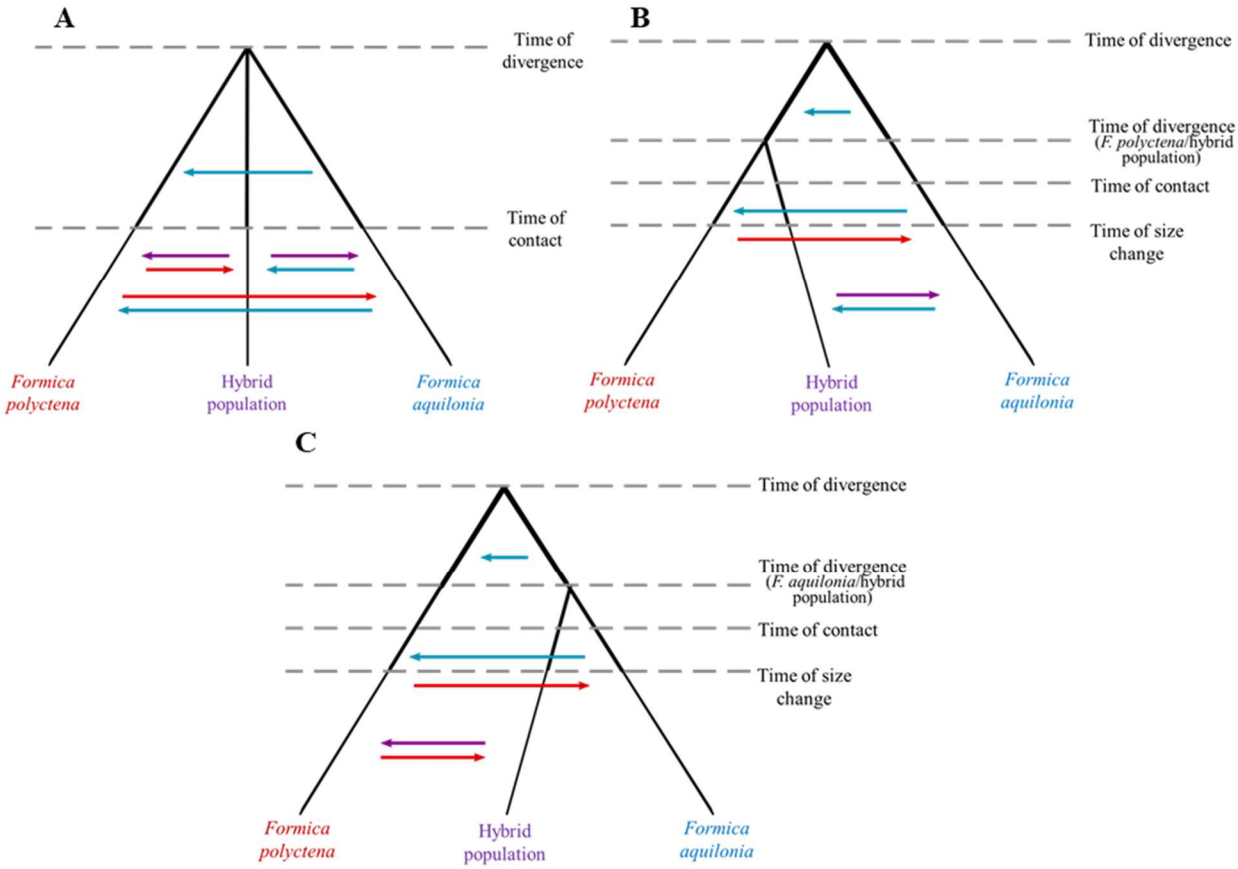


Figure 2.4 – Secondary Contact models designed to study the origin of the hybrid populations. **A** "Trifurcation": all populations diverge at the same time. **B** "((*Formica polyctena*, Hybrid population), *F. aquilonia*)": the *F. aquilonia* parental population diverges first, predating the divergence between the hybrid population and the *F. polyctena* population. **C** "((*F. aquilonia*, Hybrid population), *F. polyctena*)": the *F. polyctena* parental population diverges first, followed by the divergence between the hybrid population and the *F. aquilonia* parental population. Gene flow and changes in population size are depicted as in Figure 2.2.

The Secondary Contact models (Fig. 2.4) all consider that the hybrid population diverged from one or, when the split between the three populations is synchronous, both parental populations. There is a period of isolation after all populations have diverged from each other, after which the hybrid population

exchanges migrants with both ("Trifurcation" mode; Fig. 2.4A) or the more distantly related parental species ("(*F. polyctena*, Hybrid), *F. aquilonia*" and "(*F. aquilonia*, Hybrid), *F. polyctena*" models; Fig. 2.4B,C). From the time of the first divergence until the start of the secondary contact, migrants may only move from *F. aquilonia* into *F. polyctena*. Following the start of secondary contact, migrants are allowed to move in both directions between the parental populations. In the "Trifurcation" model (Fig. 2.4A), all three populations undergo a simultaneous size change at the time that secondary contact starts. In the "(*F. polyctena*, Hybrid), *F. aquilonia*" and "(*F. aquilonia*, Hybrid), *F. polyctena*" models, the parental populations undergo an initial size change at the time of the second divergence event, followed by a simultaneous resize for all three populations posterior to the start of secondary contact.

Our Admixture models (Fig. 2.5) consider that the hybrid population arises from an admixture event between the parental populations, where the *F. polyctena* population provides a genetic input of α into the hybrid population, while the *F. aquilonia* population inputs 1-α. As observations of the localities that are known to harbour hybrid individuals indicate that these populations may have been formed as recently as 50 years ago, both Admixture models consider that the maximum possible time of admixture is 50 generations. With our assumed generation time, the admixture events can only have happened up to 125 years ago, at most. While "Admixture" (Fig. 2.5A) considers that there is no contact between the three populations after the admixture event, "Admixture with Continuous Migration" (Fig. 2.5B) instead considers that the hybrid population continuously exchanges migrants with both parental populations since its origin until present. Both of these models consider that the parental populations undergo two size changes, with the first happening between the time of their divergence and the time of admixture, and the second happening at the same time of the admixture event.

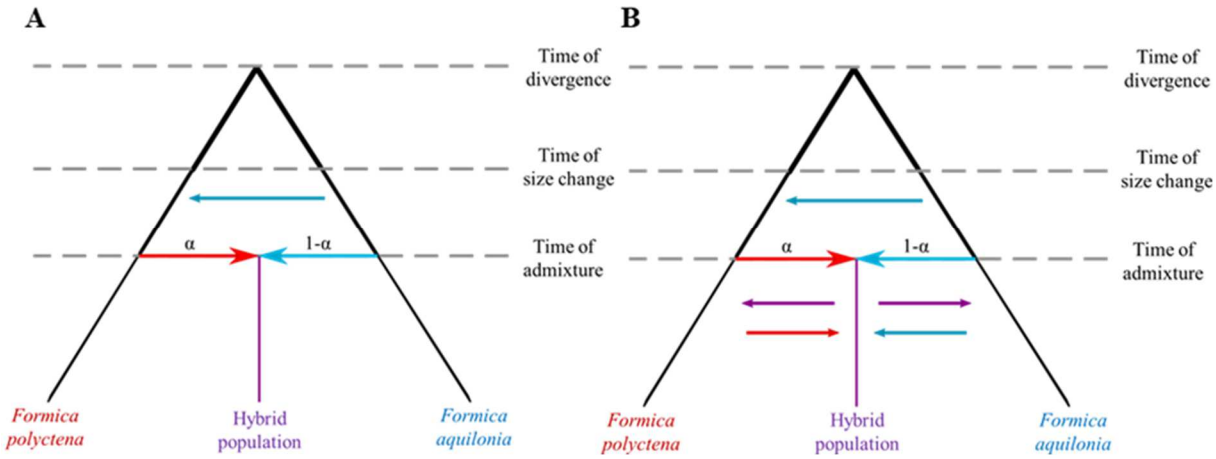


Figure 2.5 – Admixture models designed to study the origin of the hybrid populations. **A** "Admixture": the hybrid population originates from an admixture event, after the divergence of the parental populations. **B** "Admixture with Continuous Migration": after admixture, the hybrid population continuously exchanges migrants with both parental populations. Gene flow and changes in population size are depicted as in Figure 2.2.

By testing the Secondary Contact models against the Admixture models, we can more accurately infer whether the observed patterns of variation in the hybrid individuals are caused by admixture between parental *F. polyctena* and *F. aquilonia* genomes or if they are mimicked by recent divergence from one or both parental species, followed by secondary contact.

Previous studies (e.g., Beresford *et al.*, 2017) have raised the question of whether the documented populations of *F. polyctena* x *F. aquilonia* hybrid individuals have a single origin (i.e., if there was an admixture event that gave rise to an ancestral hybrid population, which then colonized the remaining geographical locations where these individuals occur by means of successive divergence events) or if each of the hybrid populations has its own independent origin (i.e., if there were multiple admixture

events happening at different points in time which gave rise to each hybrid population we sampled). Our two 4-population models (Fig. 2.6) reflect these contrasting scenarios and are appropriately named "Single Origin" and "Independent Origins". Once again, we used the same solitary individuals of the parental species sampled close to the hybrid populations as representatives of the parental populations. These models were tested a total of four times for different groups of two hybrid populations, detailed in section 2.4.2.2

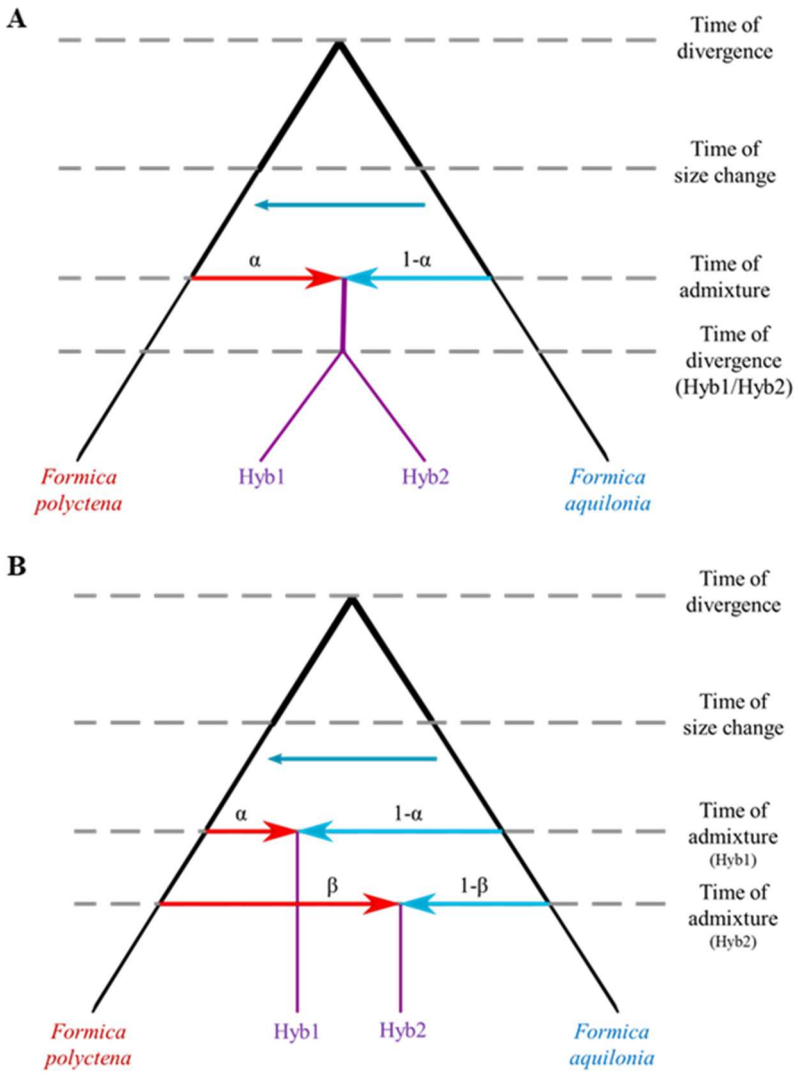


Figure 2.6 – Demographic models designed to study the origin of the hybrid populations in relation to each other. **A** "Single Origin": after the divergence between the parental populations, an admixture event originates a hybrid population that will later diverge into the remaining hybrid populations in the model. **B** "Independent Origins": after the divergence between the parental populations, there are two independent admixture events that lead to the formation of each hybrid population in the model. Gene flow and changes in population size are depicted as in Figure 2.2.

The "Single Origin" model (Fig. 2.6A) considers that, posterior to the divergence between the parental populations, an admixture event gives rise to an ancestral hybrid population. Similarl to the Admixture models, the *F. polyctena* parental population provides a proportion α of the genetic material of the hybrid population, with the *F. aquilonia* parental population providing the complementary 1-α. The ancestral hybrid population later diverges into two hybrid populations currently in existence, depicted as Hyb1 and Hyb2. In "Independent Origins" (Fig. 2.6B), we instead consider two independent admixture events at the origin of Hyb1 and Hyb2. In this model, Hyb1 receives α from *F. polyctena* and 1- α from *F. aquilonia,* and Hyb2 receives β from *F. polyctena* and 1-β from *F. aquilonia*. While the figure depicts

the incipience of Hyb1 as the first admixture event, the model does not enforce this, i.e., either of the admixture events can be the first to take place.

These models follow the "Admixture" model quite closely and, as such, also consider two separate size changes for the parental populations and exclusive migration from *F. aquilonia* into *F. polyctena* from the time of the divergence of the parental populations until the time of admixture. The admixture events cannot have happened more than 50 generations ago in either of these models.

2.4.2.2. SFS characteristics

To perform the demographic analyses detailed above, we built SFSs using data from two, three and four populations (2D-, 3D- and 4D-SFSs, respectively). As we could not polarize the SNPs and accurately infer their ancestral state, all SFSs were built using the minor allele frequency (MAF) method, and are, therefore, folded SFSs. The MAF method considers that the less frequent allele at a particular site corresponds to the "derived" state of that site.

For all SFSs, we downsampled the data to ensure that there was no missing data. To do this, a minimum sample size across all sites was determined (corresponding to the number of individuals to resample from minus the maximum number of missing data per site). Resampling the data of each individual according to the minimum sample size enabled the assembly of the SFSs with data for all sites. In all cases, individuals were resampled in windows of 50Kbp. The minimum distance between consecutive SNPs in a given block was 2 bp. The window size chosen corresponds to the distance at which we can expect sites to be considered independent or unlinked, as r^2 (a measure of LD based on the squared correlation of alleles at two loci; Hahn, 2018) reaches a plateau at this distance. To maximize the number of sites that could be kept, we resampled a lower number of individuals than those in the entire sample in each window. As the individuals selected to be resampled in each window will be the ones with higher amounts of data in that specific window, they will not necessarily be the same in all windows. This means that the SFSs will still contain information from all the individuals in our samples.

All SFSs included the number of monomorphic sites. This corresponds to number of sites whose frequency is zero in all populations in a dataset. Having the monomorphic sites information in our SFS, in conjunction with a mutation rate, allows us to scale the parameter estimates inferred by the models and obtain them in absolute terms. We estimated these numbers using the proportion of polymorphic sites in relation to the total number of callable sites of individuals in a specific dataset (*proportion*). As the polymorphic sites of a subset of individuals undergoes further modifications while the SFS is built, we must estimate how many callable sites we would be left with if they were "filtered" in the same way as the polymorphic sites. As such, we used the *proportion* and the number of sites in a dataset to estimate the number of "filtered" callable sites ($n_{callableFiltered}$). The number of monomorphic sites ($n_{monomorphic}$) was then obtained by subtracting the number of sites in each SFS ($n_{SNPsFiltered}$) from the number of filtered callable sites:

$$n_{monomorphic} = n_{callableFiltered} - n_{SNPsFiltered} \quad (2.1)$$

Where,

$$n_{callableFiltered} = \frac{n_{SNPsFiltered}}{proportion} \quad (2.2)$$

For the 3D- and 4D-SFSs, we used the proportion obtained from the dataset with the *F. polyctena* and *F. aquilonia* individuals sampled in Finland, as they were used as the parental populations.

For each 2D-SFS, we considered one population of each parental species, from which we resampled two individuals in each window. As these SFSs were used to answer our first two questions, related to the speciation history between the parental species, we compared populations non-Finnish populations, both geographically distant and near. The *F. polyctena* population in West Switzerland was compared to the *F. aquilonia* populations in Scotland and in Switzerland. The Scottish *F. aquilonia* population was also compared to the East Switzerland *F. polyctena* population. The Finnish populations, which encompassed all four individuals of each species sampled in Finland, were compared only to each other. These SFSs were also used to test the models relating to the third question, "Is there evidence for gene flow from an unsampled species to either *F. aquilonia* or *F. polyctena*?"

The 3D- and 4D-SFSs were tested with our 3- and 4-populations models, respectively, which were used to answer our final and most comprehensive question, "How did the hybrid populations originate?". In both cases, we used the single individuals each parental species sampled closed to the hybrid populations from to act as parental populations. For the hybrid populations, we resampled four individuals every window to build our SFSs. The 3D-SFSs contained information of both parental populations plus one given hybrid population, the 4D-SFSs included information of the parental populations and all hybrid populations. We analysed four different combinations of two hybrid populations, Bunkkeri and Pikkala, Långholmen W and R, Bunkkeri and Långholmen W, and Pikkala and Långholmen W.

3. Results

3.1. Hybrid populations deviate from expectations under Hardy-Weinberg Equilibrium

Mean expected heterozygosity (H_e) per population ranged from 0.123 to 0.185 (Table 3.1; Figure 3.1). The *Formica aquilonia* population in Scotland has the lowest value (0.123), while the *F. polyctena* population in Finland has the highest H_e (0.185). Mean observed heterozygosity (H_o) per population ranged from 0.103 to 0.207. Consistent with the mean expected heterozygosity, the lowest H_o (0.103) belongs to the *F. aquilonia* population in Scotland. Bunkkeri, a hybrid population, shows the highest mean observed heterozygosity (0.207).

Table 3.1 – Mean expected (H_e) and observed (H_o) heterozygosities, and mean F_{IS} per population. All values are rounded up to three decimal cases.

	Population	H_e	H_o	F_{IS}
Formica polyctena	Finland	0.185	0.169	0.087
	West Switzerland	0.134	0.123	0.082
	East Switzerland	0.119	0.123	-0.029
Hybrid populations	Bunkkeri	0.167	0.207	-0.245
	Pikkala	0.150	0.178	-0.189
	Långholmen R	0.165	0.178	-0.080
	Långholmen W	0.165	0.173	-0.047
Formica aquilonia	Switzerland	0.130	0.114	0.130
	Scotland	0.123	0.103	0.165
	Finland	0.143	0.134	0.060

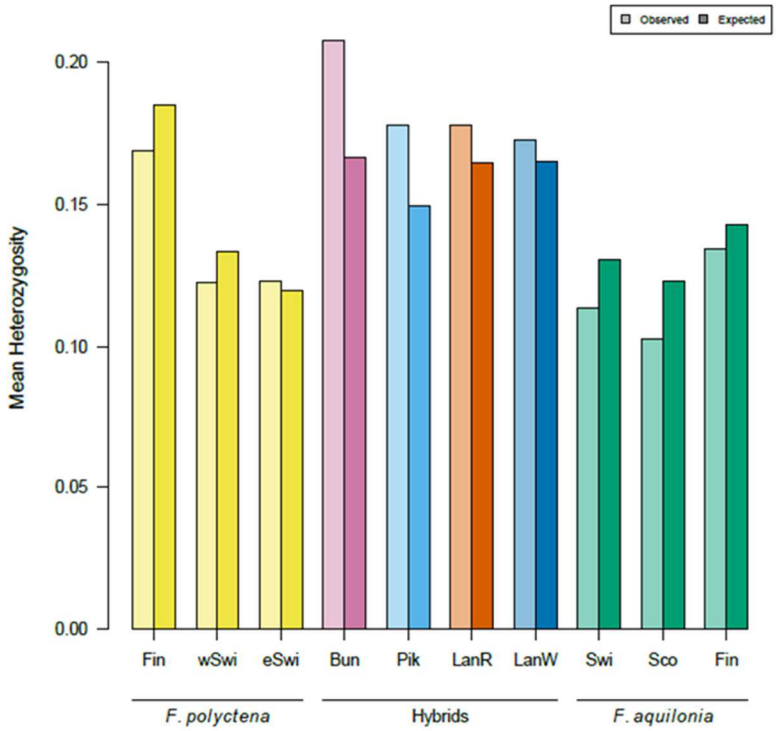


Figure 3.1 - Mean observed and expected heterozygosity of all populations under study. Abbreviations are as follows: Fin = Finland; wSwi = West Switzerland; eSwi = East Switzerland; Bun = Bunkkeri; Pik = Pikkala; LanR = Långholmen R; LanW = Långholmen W; Swi = Switzerland; Sco = Scotland

All hybrid populations have higher H_o than H_e, as well as the *F. polyctena* population in East Switzerland. This is reflected in the inbreeding coefficients (F_{IS}) of these populations, which are all negative (Table 3.1; Figure 3.2). Bunkkeri has the most negative mean F_{IS} (-0.245), while the *F.*

aquilonia population in Scotland has the most positive F_{IS} (0.165). All hybrid populations show an excess of heterozygotes, deviating from genotype frequency expectations under HWE. Furthermore, the hybrid populations, as well as the Finnish population of *F. polyctena*, are clearly different from the other populations of the parental species.

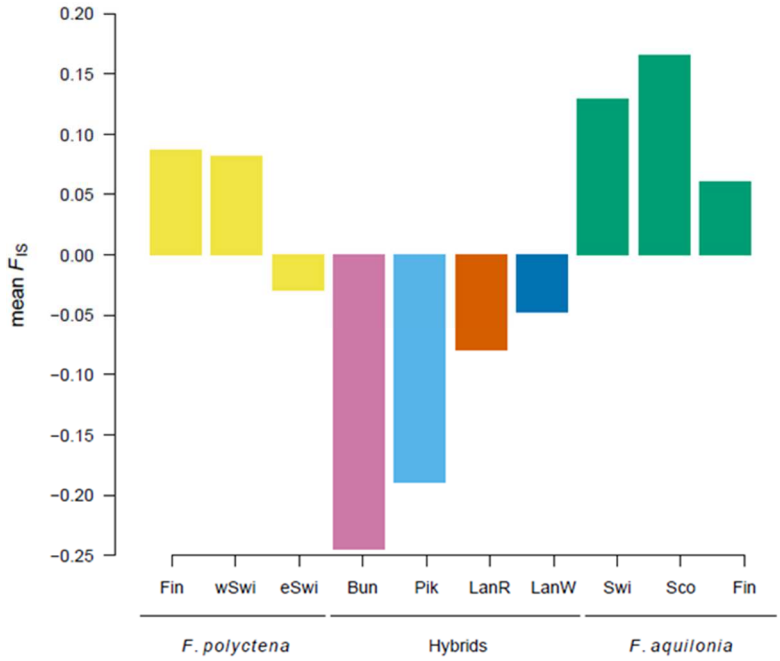


Figure 3.2 - Mean F_{IS} of all populations under study. Abbreviations are as follows: Fin = Finland; wSwi = West Switzerland; eSwi = East Switzerland; Bun = Bunkkeri; Pik = Pikkala; LanR = Långholmen R; LanW = Långholmen W; Swi = Switzerland; Sco = Scotland.

3.2. Hybrid populations are genetically intermediate between *Formica polyctena* and *Formica aquilonia*

When computing genome-wide, average pairwise differentiation indexes (F_{ST}) for all possible combinations of populations (Table 3.2; Figure 3.3), we obtained moderately high F_{ST} values (>0.1) in almost all cases. The highest value was recorded between the *F. polyctena* population in East Switzerland and the *F. aquilonia* population in Scotland (0.488), and the lowest between R and W lineages in Långholmen (-0.016). With the Weir and Cockerham (1984) estimator of F_{ST}, negative values can be taken as zero, i.e., no differentiation between populations. Average differentiation between intraspecific populations of the parental species was 0.202 for *F. aquilonia* and 0.130 for *F. polyctena*. Interspecific differentiation ranged from 0.256 to 0.497, although it tends to be lower when one or both of the populations were sampled in Finland.

Hybrid populations appear to be quite different from each other (average F_{ST} of 0.134). As pairwise F_{ST} values between hybrid populations are higher when one of the populations involved is Pikkala, it seems that Pikkala is more differentiated from Bunkkeri and Långholmen W and R than those populations are from each other. The hybrid populations seem to be more differentiated from *F. aquilonia* (average F_{ST} of 0.252 for all pairs including one hybrid population and one *F. aquilonia* population) than from *F. polyctena* (average F_{ST} of 0.222 for all pairs). Differentiation to *F. polyctena* seemed to be attenuated when the *F. polyctena* population considered in the pairwise comparison was sampled in Finland. All F_{ST} estimates are lower than 0.2 when the Finnish *F. polyctena* population is paired with a hybrid

population, whereas all pairwise F_{ST} values of comparisons between hybrid populations and the remaining *F. polyctena* populations are higher than 0.2 and go up to 0.318.

Table 3.2 - Pairwise F_{ST} values between all populations under study. All values are rounded up to three decimal cases.

		Formica polyctena				Hybrid populations				*Formica aquilonia*			
	Population	**Finland**	**West Switzerland**	**East Switzerland**	**Bunkkeri**	**Pikkala**	**Långholmen R**	**Långholmen W**	**Switzerland**	**Scotland**	**Finland**
Formica polyctena	**Finland**	-									
	West Switzerland	-	0.113								
	East Switzerland	-	0.116	0.160							
	Bunkkeri	-	0.240	0.274	0.139						
Hybrid populations	**Pikkala**	-	0.288	0.318	0.187	0.180					
	Långholmen R	-	0.224	0.259	0.135	0.187	0.125				
	Långholmen W	-	0.229	0.260	0.132	0.180	-0.016	0.128			
Formica aquilonia	**Switzerland**	-	0.445	0.481	0.249	0.298	0.253	0.249	0.283		
	Scotland	-	0.462	0.497	0.261	0.309	0.267	0.264	0.214	0.300	
	Finland	-	0.413	0.446	0.202	0.250	0.214	0.208	0.189	0.204	0.256

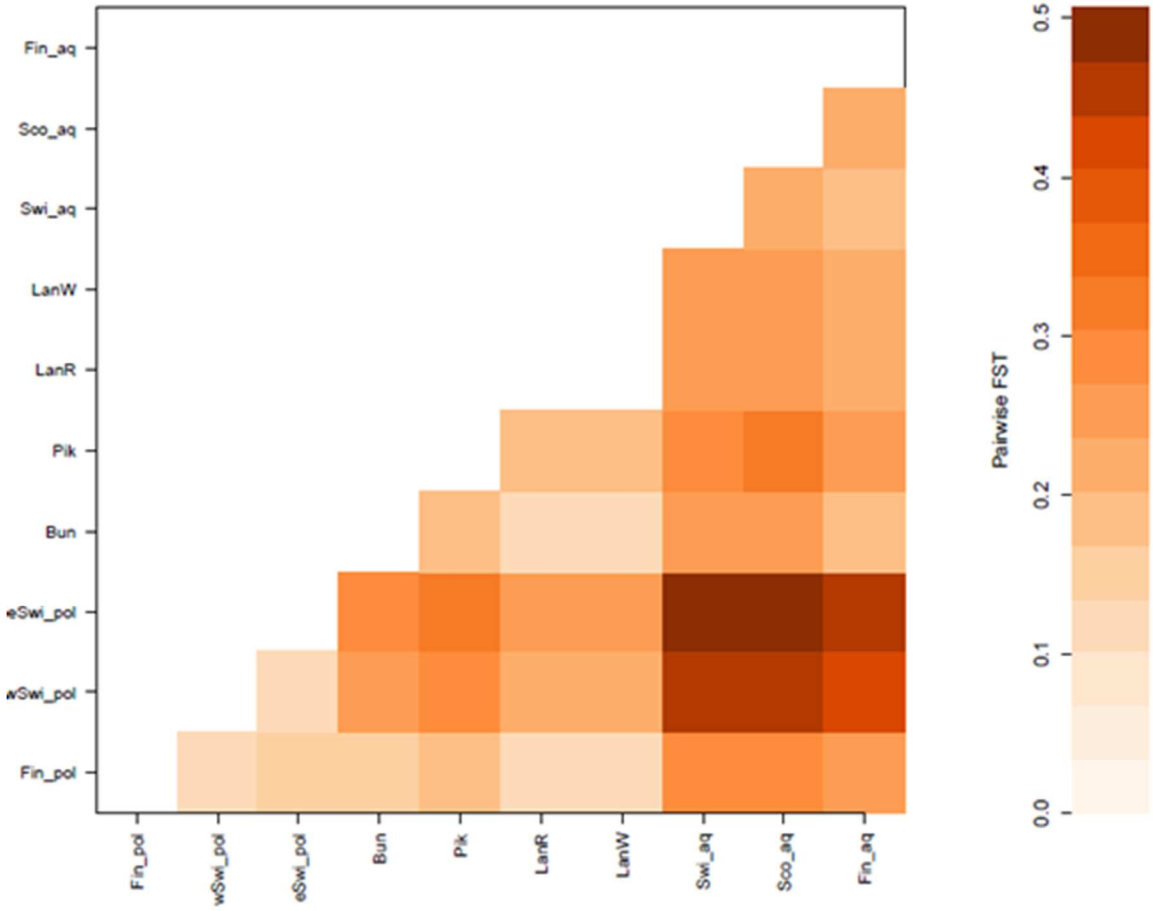


Figure 3.3 - Heat map of pairwise F_{ST} values between all populations under study. Abbreviations are as follows: Fin_pol = *F. polyctena* population in Finland; wSwi_pol = *F. polyctena* population in West Switzerland; eSwi_pol = *F. polyctena* population in East Switzerland; Bun = Bunkkeri; Pik = Pikkala; LanR = Långholmen R; LanW = Långholmen W; Swi_aq = *F. aquilonia* population in Switzerland; Sco_aq = *F. aquilonia* population in Scotland; Fin_aq = *F. aquilonia* population in Switzerland.

We employed two individual-based methods to further study the genetic population structure in our data, Principal Component Analysis (PCA) and sNMF. Tracy-Widom statistics (Supplementary Figure 1) determined that the variation explained by the first seven Principal Components (PC) produced by the PCA is statistically significant. The first three PCs are plotted against each other in Figure 3.4.

PC1, which explains ~12% of the variation in the data, clearly separates the parental populations from each other, with the *F. polyctena* populations clustered together on the left-hand side of the plot, and the *F. aquilonia* populations clustered on the right. The hybrid individuals occupy the space between the two parental clusters. It is important to note that the Finnish *F. polyctena* individuals are plotted closer to the hybrids than any other individual of either parental species. PC2 explains ~6% of the variance and reflects the differences between hybrid individuals sampled in different localities, most notably between those sampled in Pikkala and those at the other hybrid locations. Individuals from the two Långholmen lineages are clustered together and seem to be more different from Bunkkeri than they are to each other. PC3 mainly reflects the differences between individuals of the parental species sampled in Finland and those sampled in other areas. Once again, Finnish individuals of the parental species appear to be closer to the hybrid individuals.

The sNMF analysis considered two to ten possible ancestral clusters (K). Cross-entropy analysis (Supplementary Figure 2) revealed that the best value of K for our data is six, however, it is also relevant to consider the results of K=2. From the 20 repetitions done for both values of K, we chose the ancestry proportions estimated by the repetition with the lowest cross-entropy, which are shown in Figure 3.5.

When K=2 (Fig. 3.5A), individuals of both parental species cluster with each other, with the *F. polyctena* individuals grouped in the light blue cluster and with the *F. aquilonia* individuals together in the dark blue cluster. Finnish individuals of each parental species show some ancestry from the opposite ancestral cluster, up to an ancestry proportion of ~0.17. The hybrid individuals appear as a mix of ancestry from the ancestral clusters of individuals of each parental species. In all cases, hybrid individuals show a higher proportion of ancestry from the light blue ancestral cluster (*F. polyctena* individuals) than from the dark blue cluster (*F. aquilonia*) individuals. When K=6 (Fig. 3.5B), two of the reconstructed ancestral clusters are once again formed by individuals of each parental species (dark green ancestral cluster groups together all *F. polyctena* individuals, *F. aquilonia* individuals are grouped in the light green ancestral cluster). The remaining four ancestral clusters group together hybrid individuals sampled at the same location, with Bunkkeri individuals represented in light blue, Pikkala in dark blue, Långholmen R in red and Långholmen W in pink.

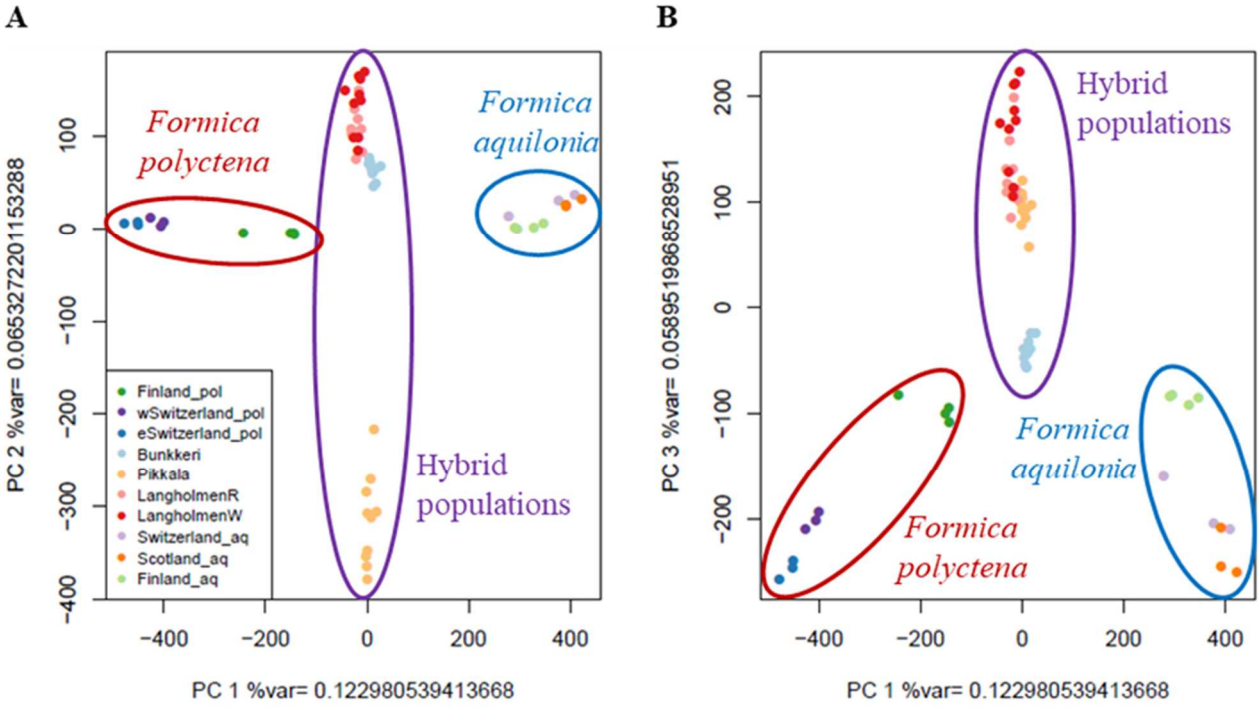


Figure 3.4 - Principal Component Analysis. Principal Components (PCs) are shown plotted against each other. Each dot represents an individual and each colour represents a population. **A** PC1 plotted against PC2. **B** PC1 plotted against PC3. Abbreviations are as follows: Finland_pol = *F. polyctena* population in Finland; wSwitzerland_pol = *F. polyctena* population in West Switzerland; eSwitzerland_pol = *F. polyctena* population in East Switzerland; Switzerland_aq = *F. aquilonia* population in Switzerland; Scotland_aq = *F. aquilonia* population in Scotland; Finland_aq = *F. aquilonia* population in Switzerland

When K=2, the Finnish populations of both parental species show ancestry of the other ancestral cluster. Particularly, Finnish *F. polyctena* individuals show higher proportions of ancestry from the *F. aquilonia* cluster (dark blue) than the proportions of ancestry Finnish *F. aquilonia* individuals show from the *F. polyctena* cluster (light blue).Individuals of both parental species sampled outside of Finland show no ancestry from the opposing cluster, except for one *F. aquilonia* individual sampled in Switzerland. When K=6, hybrid individuals show ancestry of all ancestral clusters, with varying proportions. Individuals in Bunkkeri appear to share the most ancestry with other clusters out of all hybrid populations.

These analyses suggest that the parental populations considered in this work, *F. polyctena* and *F. aquilonia*, are quite different from each other and that the hybrid individuals are genetically intermediate between the parental species.

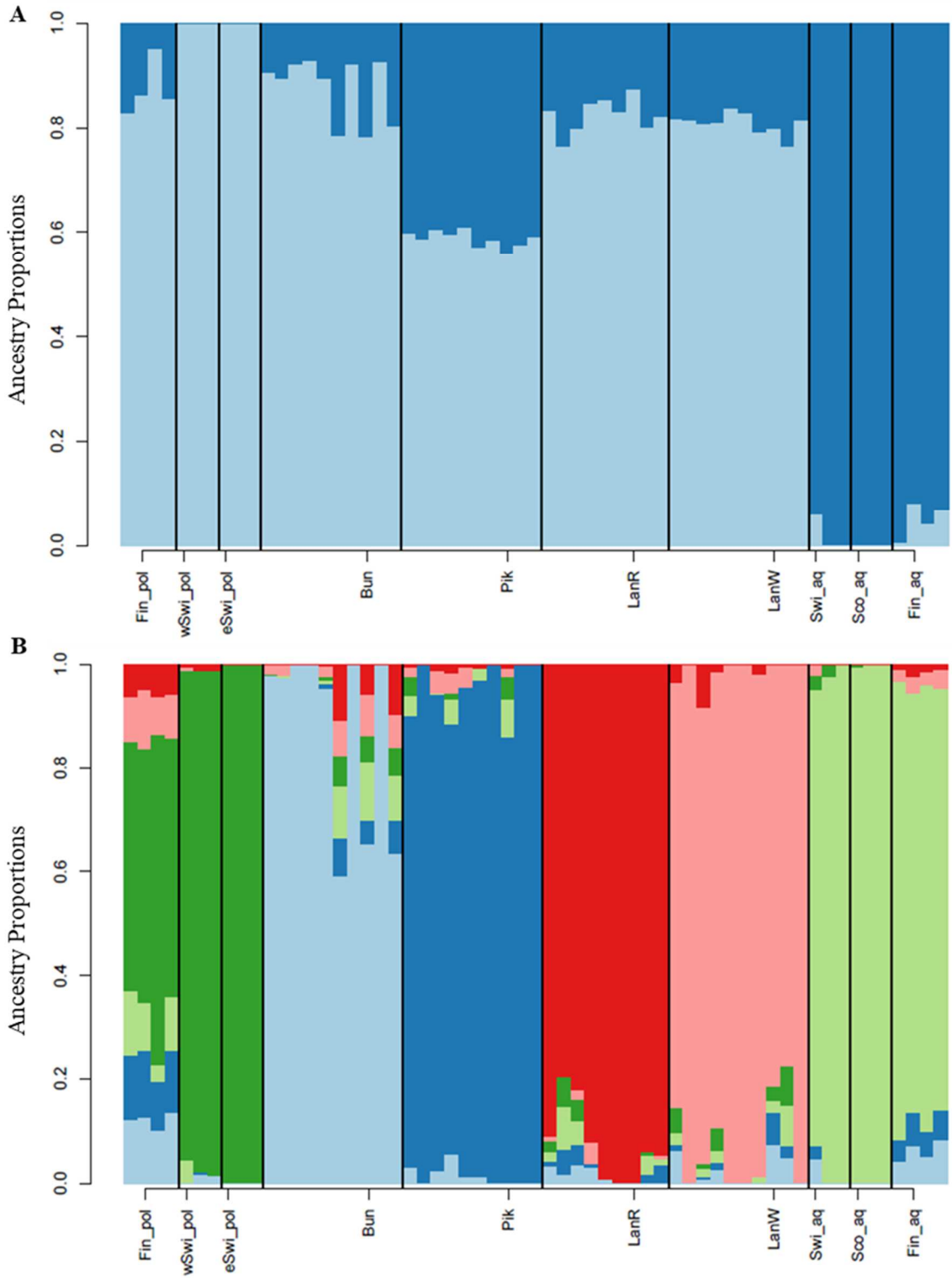


Figure 3.5 - Ancestry proportions reconstructed by sNMF for **A** K=2 and **B** K=6. Each bar corresponds to an individual and the different proportion of colours represent the probability of belonging to a specific cluster. Each population is separated by a black line. Abbreviations are as follows: Fin_pol = *F. polyctena* population in Finland; wSwi_pol = *F. polyctena* population in West Switzerland; eSwi_pol = *F. polyctena* population in East Switzerland; Bun = Bunkkeri; Pik = Pikkala; LanR = Långholmen R; LanW = Långholmen W; Swi_aq = *F. aquilonia* population in Switzerland; Sco_aq = *F. aquilonia* population in Scotland; Fin_aq = *F. aquilonia* population in Switzerland

3.3. *Formica polyctena* and *Formica aquilonia* diverged with gene flow

In order to study the speciation history between *F. polyctena* and *F. aquilonia*, we tested several models with pairs of European populations (i.e., populations sampled outside Finland). These pairs compared

the *F. polyctena* population in West Switzerland to the *F. aquilonia* populations in Scotland and Switzerland, and the *F. polyctena* population in East Switzerland to the *F. aquilonia* population in Scotland. As the datasets we used contained information on the number of monomorphic sites and we have an estimate of the mutation rate of these species, we were able to scale the parameters in a way that allows us to interpret them in an absolute manner. After testing the four models (detailed in section 2.4.2.1.1 of Material & Methods) for all the population pairs, we picked the model with the highest expected likelihood, i.e., the one that fit the data better, as the best model for each pair. This is the case for all demographic modelling results presented in this chapter.

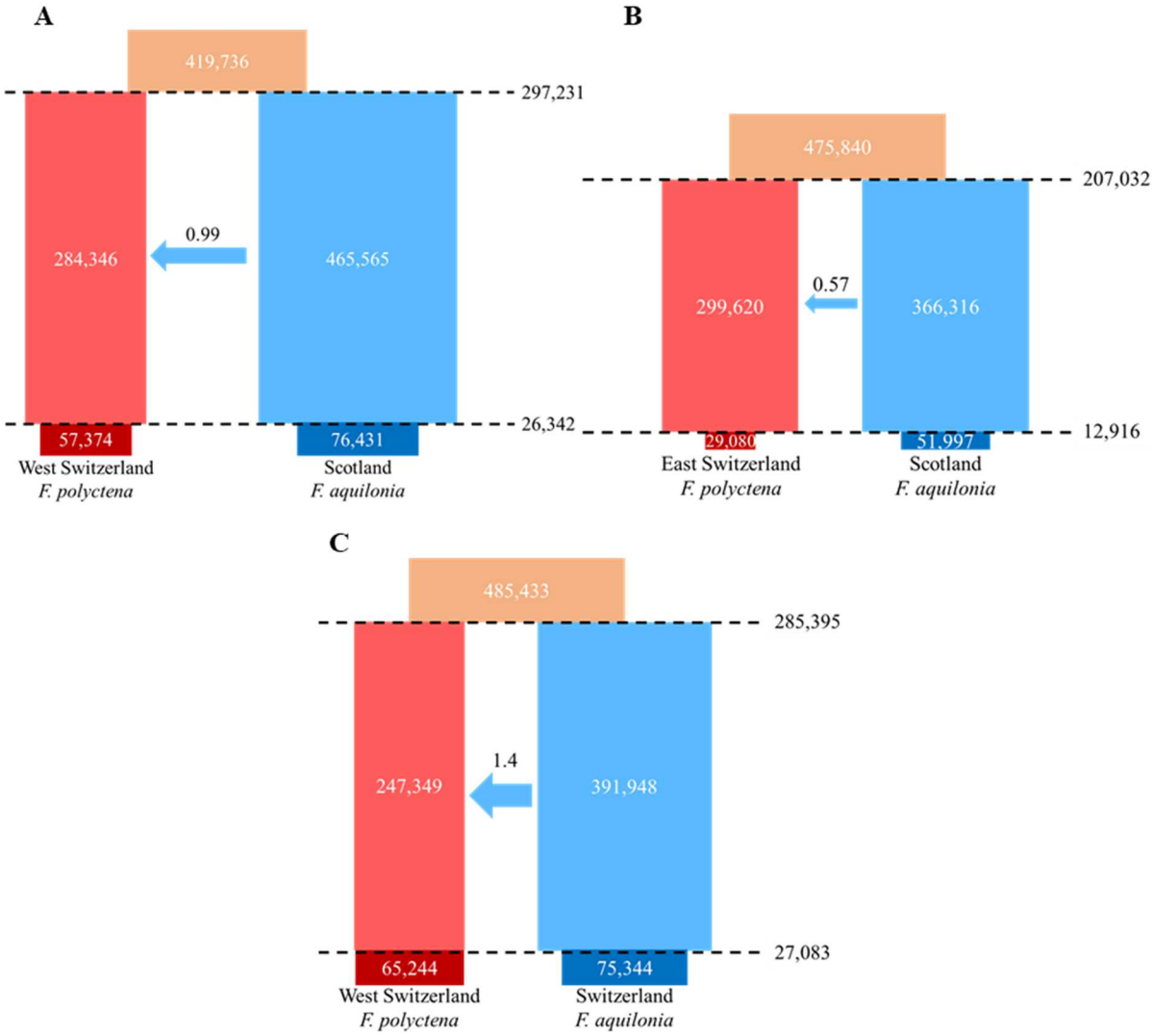


Figure 3.6 - Demographic history results for the models concerning the speciation history between *Formica polyctena* and *Formica aquilonia*. **A** Parameter estimates of the best model ("Sympatry") for the West Switzerland *F. polyctena* + Scotland *F. aquilonia* comparison. **B** Parameter estimates of the best model ("Sympatry") for the East Switzerland *F. polyctena* + Scotland *F. aquilonia* comparison. **C** Parameter estimates of the best model ("Isolation after Migragtion") for the West Switzerland *F. polyctena* + Switzerland *F. aquilonia* comparison. All times are given in number of generations and represented proportionally to each other across panels, as the time of divergence in panel A was taken as reference. All effective sizes are given in number of haploids. Sizes at a given time are represented proportionally to each other across panels, with the *F. polyctena* sizes in panel A serving as reference (i.e., all recent sizes are proportional to each other but not to ancestral sizes, while all ancestral sizes are proportional to each other but not to recent sizes). Arrows indicate the number of migrants per generation, their size is representative of this value. The direction and colour of the arrows are indicative of the direction of the gene flow.

The "Sympatry" model was the best fit for the West Switzerland *F. polyctena* x Scotland *F. aquilonia* and East Switzerland *F. polyctena* x Scotland *F. aquilonia* comparisons (Fig. 3.6A,B; Supplementary Tables 2 and 3 for parameter estimates obtained with all models for both comparisons). We found

"Isolation after Migration" to be the best model for the West Switzerland *F. polyctena* x Switzerland *F. aquilonia* comparison (Fig. 3.6C; Supplementary Table 4 for parameter estimates obtained with all models). The best parameter estimates for each population comparison can be found in Figure 3.6, in their respective panels.

The time at which the populations of each species diverged is consistent across the different population comparisons. The smallest observed estimate for the divergence time, 207,032 generations, was obtained for the comparison between East Switzerland *F. polyctena* and Scotland *F. aquilonia*, and the largest estimate of 297,231 generations was obtained for the West Switzerland *F. polyctena* x Scotland *F. aquilonia* population pair. Assuming a generation time of 2.5 years, the estimates for the divergence between these species range from 517,580 to 743,077.5 years ago, depending on the population pair.

The ancestral population of both species is consistently estimated to have an effective size between 400,000 and 500,000 haploid individuals across population comparisons. After the divergence of the species, *F. aquilonia* is consistently estimated, across population comparisons, to have a larger N_e than *F. polyctena* throughout their history. The models consider that both populations undergo simultaneous size changes, and our results for all comparisons indicate that both species suffer contractions at the time of the size change. The best estimates we obtained for the time of the size change ranged from 12,916 generations, for the East Switzerland *F. polyctena* x Scotland *F. aquilonia* comparison, to 27,083 generations, for the West Switzerland *F. polyctena* x Switzerland *F. aquilonia* comparison. These size changes are estimated to have happened 32,290 to 67,708 years ago.

The best models indicate that *F. polyctena* and *F. aquilonia* diverged with gene flow. This gene flow is very asymmetrical across population comparisons, with migrants moving exclusively from *F. aquilonia* into *F. polyctena*. Our estimates of the number of immigrants (*2Nm*) moving in this manner every generation ranged from 0.57, for the comparison between East Switzerland *F. polyctena* and Scotland *F. aquilonia*, to 1.4, for the West Switzerland *F. polyctena* x Switzerland *F. aquilonia*.

3.4. Finnish populations reveal the same speciation history, with bidirectional gene flow, between *Formica polyctena* and *Formica aquilonia*

To investigate whether the history of divergence between these species is different in Finland due to, for example, the occurrence of hybridization in this area, we tested the same models as before with a dataset comparing the populations of both species sampled in Finland.

Similarly to what we obtained with pairs of European (i.e., non-Finnish) populations, the "Sympatry" model is the best fit for the Finnish *F. polyctena* and *F. aquilonia* populations (Fig. 3.7; Supplementary Table 5 for parameter estimates obtained with all models). In other words, this comparison also supports a scenario of divergence with gene flow for *F. polyctena* and *F. aquilonia*. The time of divergence between these populations, estimated as 224,481 generations (561,202.5 years), is in-line with previous estimates. The size of the ancestral population of both populations in the model is quite comparable to the estimates obtained for the European comparisons, and the ancestral populations of both species follow the previous trend with larger estimates for *F. aquilonia* than *F. polyctena*. However, at the time when the population sizes change, the size of Finland *F. polyctena* increases, i.e., this population expands while we consistently saw *F. polyctena* contracting in the previous analyses. At the time of the size change, Finland *F. aquilonia* still contracts, but both Finnish populations are estimated to be bigger than other conspecific populations at a more recent time. The time of size change is estimated to be older than what we saw previously, being placed at 39,277 generations, or 98,193 years.

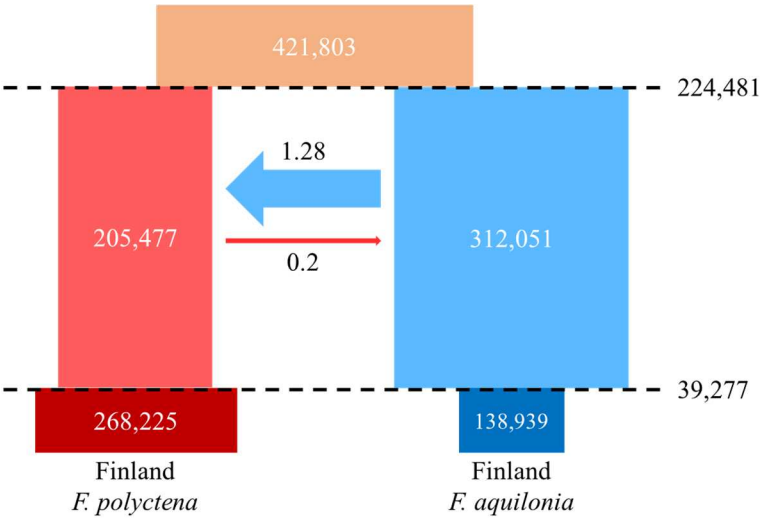


Figure 3.7 – Best demographic history for the Finnish populations of *Formica polyctena* and *Formica aquilonia*. Results are displayed as in Figure 3.6.

Similar to what we saw in the previous set of results, there is considerable gene flow from *F. aquilonia* into *F. polyctena*, with 1.4 migrants moving in this manner every generation. However, unlike what we saw with the European comparisons, there is also gene flow from *F. polyctena* into *F. aquilonia* in Finland, at a rate of 0.2 migrants every generation.

3.5. Past gene flow between *Formica polyctena* and *Formica aquilonia* cannot be explained by gene flow from unsampled sister species

To explore the possibility that the observed pattern of gene flow between these species is caused by migration from an unsampled, more closely related species into *F. polyctena* or *F. aquilonia*, we tested two models that included unsampled ("ghost") populations. Details of these models can be found in Section 2.4.2.1.2 of the Material & Methods. These models were tested for both European and Finnish comparisons (Supplementary Tables 6 and 7 for the parameter estimates obtained with both models for all population pairs).

The expected likelihood of these models with ghost admixture is lower than all other models that considered migration between the sampled species to be possible in some way. Importantly, models with ghost admixture are worse than the best model of each population pair. As such, possible unaccounted migration from a more closely related species into either *F. polyctena* or *F. aquilonia* cannot explain the observed pattern of asymmetric, direct migration between these species, mostly from *F. aquilonia* into *F. polyctena*. However, the "(*F. polyctena*, Ghost), *F. aquilonia*" consistently estimates considerable amounts of migrants moving from the unsampled population into *F. polyctena* for almost all pairs of populations, with the exception of the East Switzerland *F. polyctena* x Scotland *F. aquilonia* comparison.

3.6. Hybrid populations arose from admixture between *Formica polyctena* and *Formica aquilonia*

To investigate the origin of the hybrid populations, we tested several models detailed in section 2.4.2.1.3 of the Materials & Methods. We used the samples collected in Southern mainland Finland as sole representatives of the parental populations due to their proximity to the hybrid populations and their quality as representatives of the parental species.

Since we tested models that alternatively consider that the hybrid populations result from either secondary contact or admixture between the parental populations, we can more confidently assert that admixture between *F. polyctena* and *F. aquilonia* is at the origin of the hybrid populations. This is because we can objectively say that both the models that contain an admixture event as the origin of the hybrid populations are a better fit to our data than any of the models with secondary contact. In other words, the models with admixture have higher expected likelihoods in all cases. Furthermore, the simple "Admixture" model, with no backcrossing with the parentals and no post-admixture migration between the parental populations, is the best fit for all hybrid populations (Fig. 3.8; Supplementary Tables 8-15 for the parameter estimates obtained with all models for all hybrid populations).

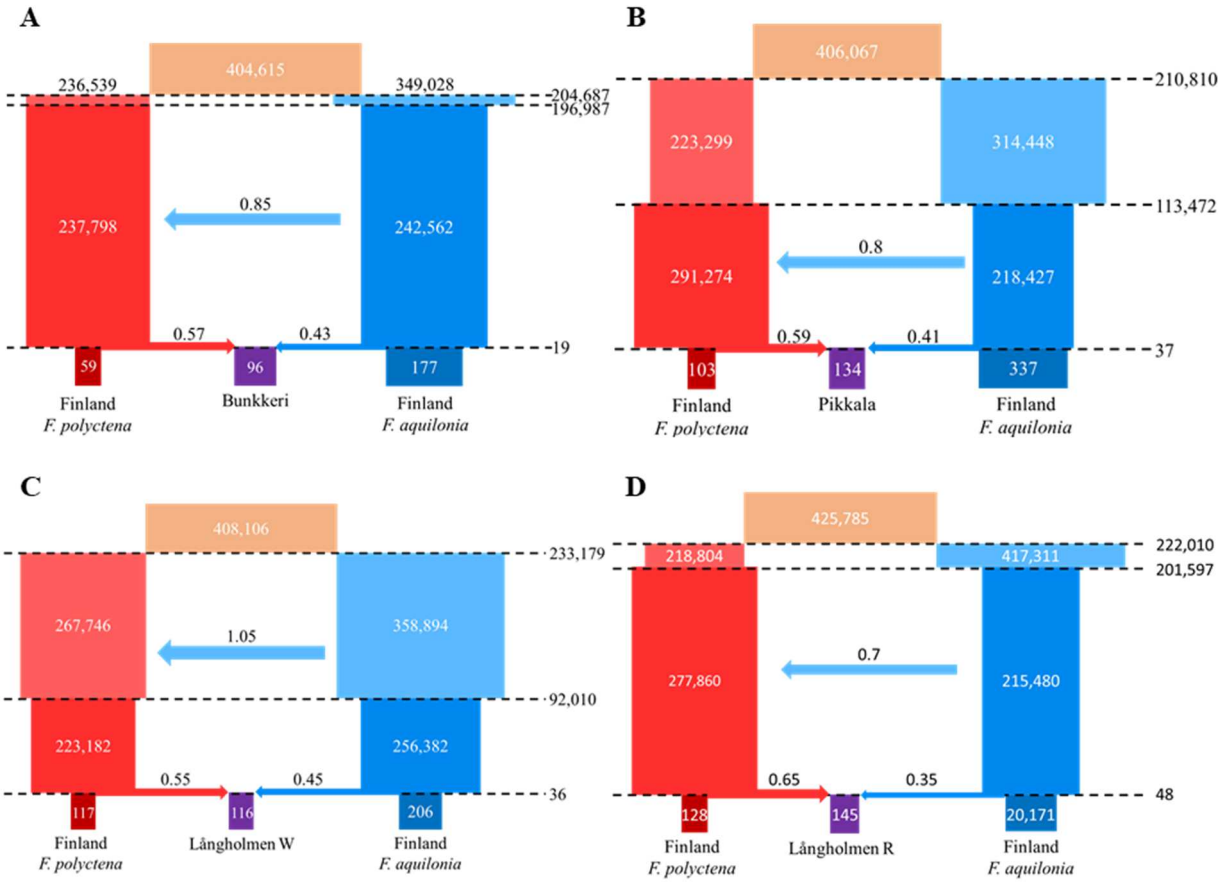


Figure 3.8 – Parameter estimates of the "Admixture" model for **A** Bunkkeri, **B** Pikkala, **C** Långholmen W, and **D** Långholmen R, and their parental populations. Results are displayed as in Figure 3.6, except for the recent size of *Formica aquilonia* in panel D.

It is a common trend across the results of these analyses that the recent *F. aquilonia* parental population is the largest of the three, while the recent *F. polyctena* population is the smallest. The hybrid populations tend to be somewhat bigger than the recent *F. polyctena* parental population. It is also consistently estimated across analyses that *F. polyctena* contributed more genetic material into the hybrid populations than *F. aquilonia*. When the hybrid population considered is Bunkkeri, Pikkala or Långholmen W, the model estimates that *F. polyctena* contributed 55-59% of the genetic material of these hybrid populations. When the hybrid population is Långholmen R, *F. polyctena* is estimated to have contributed 65% of the genetic material of this population.

Långholmen R (Fig. 3.8D) is estimated to be the oldest of the hybrid populations, with the admixture event from which it originated being estimated to have happened 120 years ago. The time at which Pikkala (Fig. 3.8B) and Långholmen W (Fig. 3.8C) originated differs by one generation, with their

respective origins being estimated to have happened 93 years ago for Pikkala, and 90 years ago for Långholmen W. The time at which Bunkkeri (Fig. 3.8A) originated is estimated to be 48 years ago, suggesting that Bunkkeri is the youngest of the hybrid populations.

3.7. W and R lineages in Långholmen share the same origin

In order to explore whether the hybrid populations were all formed through independent admixture events or if their origin is shared, implying also shared ancestry between them, we tested two models where sets of two hybrid populations were considered along with their parental populations. Details of these models can be found in Section 2.4.2.1.3 of the Material & Methods. We considered four groups of hybrid populations, Långholmen W and R (Fig. 3.9), Bunkkeri and Långholmen W (Fig. 3.10), Pikkala and Långholmen W (Fig. 3.11), and Bunkkeri and Pikkala (Fig. 3.12).

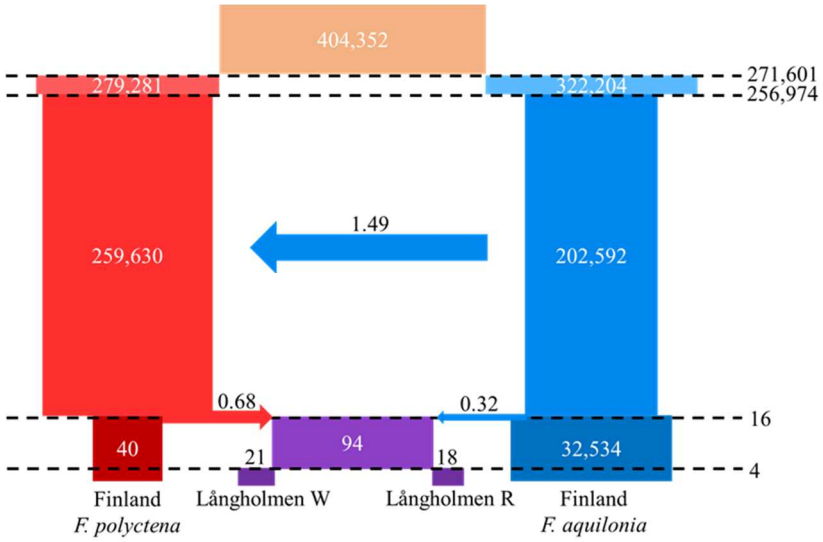


Figure 3.9 - Best parameter estimates of the "Single Origin" model for the dataset with the Långholmen W and R hybrid populations, as well as their parental populations. Results are displayed as in Figure 3.6.

We found that the "Single Origin" model was the best fit in all cases (Supplementary Tables 16 and 17 for results of both models for all hybrid population groups). However, the comparison between the two lineages in Långholmen is the only instance where this model was distinctly better than the "Independent Origins" model. Adding to the observation that the expected likelihood for "Independent Origins" is over 2,000 log units worse than the expected likelihood for the "Single Origin" model, the parameter estimates inferred by the "Single Origin" model strongly suggest that there was a single admixture event between *F. polyctena* and *F. aquilonia* 40 years ago, followed by three decades of shared ancestry. This would mean these populations spent 75% of their existence together, having very recently separated into independent populations. The results of this analysis imply that *F. polyctena* contributed 68% of the genetic material of the ancestral population of Långholmen W and R. We are not able to paint such a clear picture for the other groups of hybrid populations. For the remaining cases, the expected likelihood of the "Independent Origins" model is worse than that of the "Single Origin" model by only ~100 log units and the parameter estimates of the best model point towards very short periods of shared ancestry prior to the split of the ancestral hybrid population into the populations we see now. For each pair, "Independent Origins" estimates that the independent admixture events happened nearly simultaneously in most cases. These estimates agree with those of the "Single Origin" model towards the time of the admixture event that originates the ancestral hybrid population and the time of the subsequent divergence into the present hybrid populations.

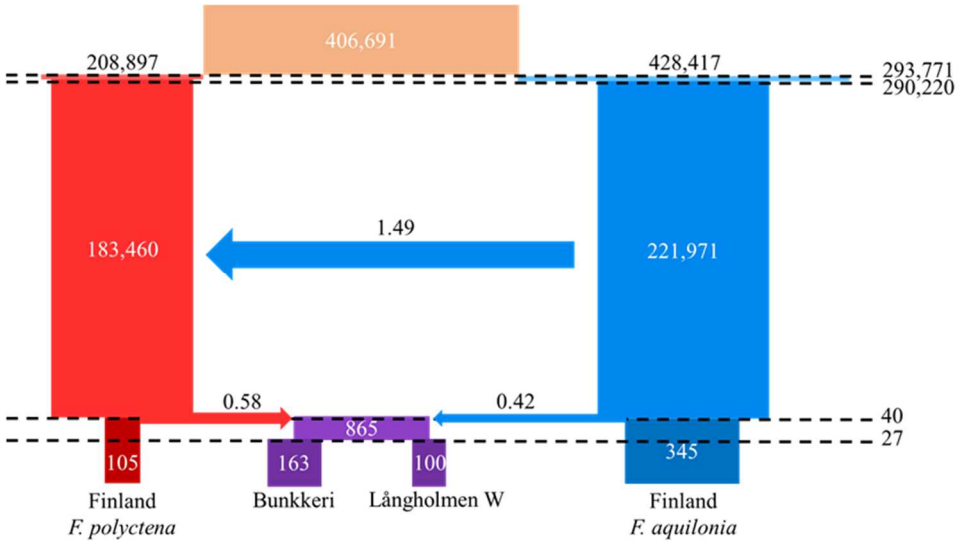


Figure 3.10 - Best parameter estimates of the "Single Origin" model for the dataset with the Bunkkeri and Långholmen W hybrid populations, as well as their parental populations. Results are displayed as in Figure 3.6.

Additionally, the sizes of the ancestral hybrid populations in the "Single Origin" model are consistently estimated to be quite high, compared to the recent sizes of the hybrid populations. When lineages cannot coalesce within a population, that is reflected in larger effective population sizes. As fastsimcoal2 implements a coalescent-based method, it is likely that these sizes are inflated due to the lack of coalescent events between lineages from the two hybrid populations in the ancestral population, again consistent with an independent origin. The opposite happens with the ancestral population of the Långholmen W and R populations, which is estimated to have a size of 94 haploid individuals. This indicates that the lineages of the hybrid populations involved coalesce with each other in the ancestral hybrid population. As such, this supports the hypothesis that the W and R lineages in Långholmen share a common origin, followed by a considerable period of shared ancestry.

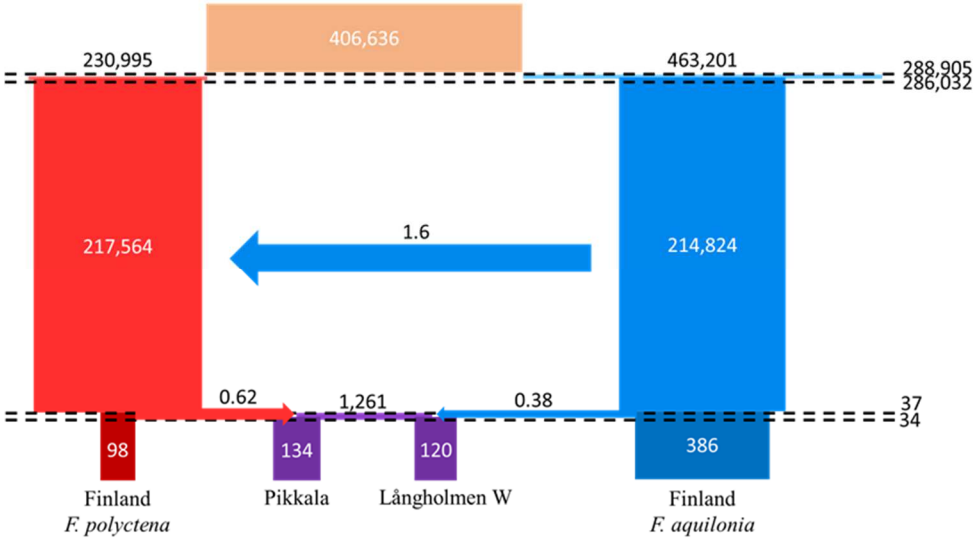


Figure 3.11 - Best parameter estimates of the "Single Origin" model for the dataset with the Pikkala and Långholmen W hybrid populations, as well as their parental populations. Results are displayed as in Figure 3.6.

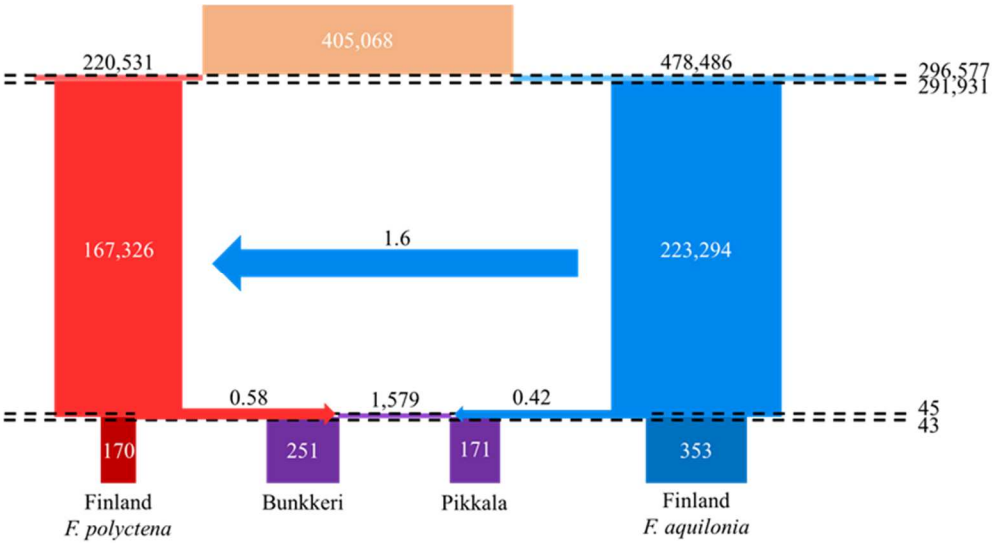


Figure 3.12 - Best parameter estimates of the "Single Origin" model for the dataset with the Bunkkeri and Pikkala hybrid populations, as well as their parental populations. Results are displayed as in Figure 3.6.

4. Discussion

We used whole-genome genomic data to study speciation and hybridization between two wood ant species, *Formica polyctena* and *F. aquilonia*. We found that *F. polyctena* and *F. aquilonia* diverged with continuous asymmetric gene flow. Our results support the hypothesis that the putative *F. polyctena* x *F. aquilonia* hybrid individuals result from admixture between these species, and suggest that the two lineages extant in the Långholmen population result from the same admixture event.

4.1. What is the speciation history between *Formica polyctena* and *Formica aquilonia?*

Previous studies on the speciation of *rufa* group ants of the *Formica* genus have implemented phylogenetic approaches and used predominantly mitochondrial and microsatellite markers (e.g., Goropashnaya *et al.*, 2004, 2007; Goropashnaya, Fedorov and Pamilo, 2004). This project marks the first instance where speciation within the *rufa* group species is studied using a large number of markers sampled across the entire genome and where estimates of important demographic parameters are obtained for both species considered, *F. polyctena* and *F. aquilonia*. Furthermore, the approach implemented here compared several populations of these species that were sampled throughout Europe. Remarkably, we inferred the same history with multiple, distinct, pairs of populations, which affords reliability to our results.

The results of our demographic analyses concerning the speciation history between *Formica polyctena* and *F. aquilonia* indicate that the divergence between these species is estimated to have happened between 517,580 to 743,078 years ago, depending on the population pair considered (Fig. 3.6). This means that the divergence likely occurred in the Pleistocene, assuming a generation time of 2.5 years. The diversification of the Formicidae family is thought to have started 110-115 million years ago (Grimaldi and Agosti, 2000), and the emergence of the Formicinae subfamily has been dated to 104-117 million years ago (Blaimer *et al.*, 2015). According to Goropashnaya, Fedorov and Pamilo (2004), the *rufa* group includes eight species, *F. rufa*, *F. polyctena*, *F. aquilonia*, *F. lugubris*, *F. paralugubris*, *F. pratensis*, *F. frontalis* and *F. truncorum*. The *F. rufa* and *F. polyctena* species form a basal monophyletic clade estimated to have separated from the remaining species in the *rufa* group 490 thousand years ago (Goropashnaya, Fedorov and Pamilo, 2004; Goropashnaya *et al.*, 2012). As *F. polyctena* and *F. aquilonia* are part of different clades, we would expect their most recent common ancestor to have existed at a time prior to the split of the *F. rufa/F. polyctena* clade from the remaining members of the *rufa* group. As such, our estimated time of divergence between *F. polyctena* and *F. aquilonia* precedes the separation of the *F. rufa*/*F. polyctena* clade from the remaining *rufa* species, which is in agreement with previous estimates of divergence in the *rufa* group.

The ancestral population of *F. polyctena* and *F. aquilonia* is inferred to have had an effective size (N_e) of 419,736 to 485,433 haploid individuals. After the divergence of these species, *F. aquilonia* is consistently inferred to have a larger ancestral N_e than *F. polyctena*. Both species are estimated to have suffered contractions 32,290 to 67,708 years ago. The sizes of both populations are inferred to have considerably decreased, with *F. aquilonia* still being consistently estimated to be larger than *F. polyctena*. Due to the supercolonial nature of *F. polyctena* and *F. aquilonia* populations, it is likely that the species follow the dynamics of a metapopulation (i.e., a population subdivided into many separate demes that may exchange genes, become extinct or recolonized after extinction; Wakeley and Aliacar, 2001). In metapopulations, coalescence of lineages within the same deme is expected to be faster than between lineages in different demes (Wakeley, 2004). We suspect that our effective size inferences may be inflated due to the existence of many lineages in the populations that cannot coalesce with each other, mimicking a large population.

The analysis of present-day populations of *F. polyctena* and *F. aquilonia* revealed high inter- and intraspecific differentiation. While conspecific individuals are clearly more similar to each other than to heterospecific individuals, we still see significant differentiation between populations of the same species. The estimates of heterozygosity we obtained for populations of *F. polyctena* and *F. aquilonia* are quite low, especially when compared to previous studies that characterized heterozygosity in *Formica* species. These previous studies often used small numbers of microsatellite and mitochondrial loci to estimate genetic diversity (Chapuisat, 1996; Goropashnaya, Seppä and Pamilo, 2001; Gyllenstrand, Gertsch and Pamilo, 2002; Seppä *et al.*, 2012), obtaining estimates as high as 0.75 (Chapuisat, Bocherens and Rosset, 2004). Evidently, our estimates of mean heterozygosity values for *F. polyctena* and *F. aquilonia* are much lower, however, we obtained these estimates using whole-genome data, with over two million SNP sites.

Our results place the timing of the size contraction of these species in the last glacial period, which lasted from circa 115,000 to 11,700 years ago. While not much is known about the phylogeographical structure of these species, both have previously been suggested to have suffered bottlenecks while surviving glaciation in suitable forest refugia, subsequently colonizing most of Eurasia (Goropashnaya, Fedorov and Pamilo, 2004). The effective size contractions inferred by our demographic analyses could be attributed to the bottlenecks that took place while the species were trapped in the refugia, followed by possible founder effects when both species expanded and colonized their remaining territory.

Our analyses strongly suggest that *F. polyctena* and *F. aquilonia* diverged with gene flow. However, our results show some discordance as to the manner in which this gene flow takes place. The "Sympatry" model, which implements a scenario of divergence with continuous gene flow, is the best fit for two out of three pairs of populations. For the dissident pair, the comparison between West Switzerland *F. polyctena* and Switzerland *F. aquilonia*, "Isolation after Migration" is the best scenario. Even so, this model estimates that the populations became isolated only 67,708 years before present time, meaning that these populations still spent over 640,000 years experiencing gene flow. While it would seem unusual that we found the *F. polyctena* population in West Switzerland to have diverged with continuous gene flow from the *F. aquilonia* population in Scotland, given that *F. polyctena* does not occur in Scotland, the demographic history inferred with these populations reflects the overall history of the species and not of these populations themselves. Here, we show evidence for interspecific gene flow between two species of the *rufa* group, which has also been described for other *Formica* species (e.g., Purcell *et al.*, 2016). Many other ant species are also known to engage in interspecific gene flow (Feldhaar, Foitzik and Heinze, 2008), with both sister (e.g., Seifert, 2019) and non-sister species (e.g., Steiner *et al.*, 2011).

It is known that many *Formica* species of the *rufa* group retain the ability to interbreed and produce viable offspring, with 56% of these species hybridizing with varying frequency (Seifert and Goropashnaya, 2004). As such, it is not surprising that we found evidence for gene flow between *F. polyctena* and *F. aquilonia*. However, one of our most unexpected results is that the gene flow between *F. polyctena* and *F. aquilonia* is consistently inferred to be asymmetrical, with only *F. aquilonia* genes flowing into *F. polyctena*. It is possible that prezygotic isolation mechanisms are stronger for *F. aquilonia* than for *F. polyctena*. Preliminary mate-choice experiment results suggest that, in Southern Finland, *F. aquilonia* individuals are more stringent when it comes to selecting a mate than the hybrid *F. aquilonia* × *F. polyctena* individuals found in this area (Beresford, Ferkinstad *et al.*, unpublished). If the lower rigour in mate-choice displayed by hybrids is due to having an intermediate phenotype between those of the parental species, then it could be assumed that mate-choice is less strict for *F. polyctena*, facilitating the movement of *F. aquilonia* genes into *F. polyctena*, rather than in the reverse direction. Mate-choice experiments would have to be conducted with *F. polyctena* and *F. aquilonia*

individuals collected across Europe to verify this hypothesis. While our inferences are obtained with sites spread across the entire genome, and are, therefore, more likely to reflect past demographic events, another alternative for the asymmetry in gene flow levels may be natural selection. It has been shown that *F. aquilonia* is more resistant to cold than *F. polyctena* (Martin-Roy, Nygård *et al.*, submitted), as such *F. aquilonia* alleles introgressed into *F. polyctena* may allow this species to perform better when temperatures are lower. If the *F. aquilonia* alleles introgressed into *F. polyctena* grant higher tolerance to low temperatures and are, therefore, beneficial for *F. polyctena* individuals, it is possible that they have been maintained in the populations throughout the history of *F. polyctena*. The scope of this project does not afford us the opportunity to offer any kind of support to this hypothesis, however, simulation work focused on the expected loss of neutral introgressed alleles in populations of these species could help elucidate this matter.

4.2. Is the history of divergence between *Formica polyctena* and *Formica aquilonia* different in Finland compared to Europe?

Hybridization between *F. polyctena* and *F. aquilonia* has already been characterized in Southern Finland (Rosengren, 1977; Sorvari, 2006; Kulmuni, Seifert and Pamilo, 2010), something that is known to happen often at the edge of the distribution of a species (Pfennig, Kelly and Pierce, 2016). Therefore, it is not unreasonable to consider that inferences of the demographic history between *F. polyctena* and *F. aquilonia* populations known to meet and produce hybrids in Southern Finland may deviate from the underlying speciation history.

The present populations of *F. polyctena* and *F. aquilonia* sampled in Finland are clearly different from the remaining non-Finnish populations of their respective species. We saw that interspecific differentiation is fairly reduced when we compare the Finnish populations. These populations were inferred to possess reconstructed ancestry from hybrid individuals sampled at different locations, and are more genetically similar to our putative hybrid populations than the other non-Finnish populations. In addition, they are more genetically diverse than their conspecific populations sampled outside Finland. Particularly, Finland *F. polyctena* has the highest genetic variability out of all the populations we sampled. This agrees with the demographic inference results, which indicate that the *F. polyctena* effective population size in Finland is the largest of the *F. polyctena* populations we sampled. These results point towards *F. aquilonia* and *F. polyctena* experiencing more gene flow in Finland than in other sampled locations. *F. polyctena* is thought to have colonized Finland after *F. aquilonia* had already become established. Theoretical work by Currat *et al.* (2008) demonstrated that, when a species expands its range and colonizes new territory, there is substantial introgression of neutral alleles from the established species into the colonizing species. Applied to our situation, this would mean that the dispersers colonizing Southern Finland would have been genetically enriched by alleles introgressed from the previously existing *F. aquilonia* gene pool, increasing the genetic diversity in the Finnish *F. polyctena* population and inflating its N_e.

The speciation history inferred with the Finnish populations of *F. polyctena* and *F. aquilonia* fits in with the overall history as inferred with other European populations. The divergence between these populations is estimated to have happened 561,203 years ago, which is very consistent and comparable to estimates obtained for pairs of non-Finnish populations. Effective sizes estimated for the ancestral population of the Finnish samples and for the ancestral populations of each species are also consistent with results obtained with the other populations. These populations are inferred to have diverged with continuous gene flow, with a considerable amount of migrants moving from *F. aquilonia* into *F. polyctena*. As such, the occurrence of hybridization between Finnish populations of *F. polyctena* and *F.*

aquilonia in Southern Finland does not seem to distort the "bigger picture", the overall speciation history.

However, there are some very noticeable discrepancies between the speciation history between *F. polyctena* and *F. aquilonia* and the history of their Finnish populations. The first of these is the time at which the size of these populations changed. These populations are inferred to have suffered size changes approximately 100,000 years ago, at an older time than any of the other populations. While we consistently saw both populations contract, we found that Finnish *F. polyctena* actually expands at the same time that Finnish *F. aquilonia* contracts. This leads to the second discrepancy, as we now see the effective size of Finnish *F. polyctena* increasing to 268,225 haploid individuals. This disagrees with the previous tendency for *F. aquilonia* to have a larger effective size than *F. polyctena* at more recent times. Most interestingly, the pattern of migration between the Finnish populations of *F. polyctena* and *F. aquilonia* differs from what we found between non-Finnish populations. Alternatively, this increase in N_e may reflect an increase in the immigration into *F. polyctena*. As the model assumes a constant migration rate *m* through time, changes in N_e will affect the average number of immigrants (i.e., the scaled immigration rate *2Nm*). Thus, the increase in N_e might reflect an increase in the immigration rate. While we previously only inferred migration from *F. aquilonia* into *F. polyctena*, we also found evidence for the movement of lineages from *F. polyctena* into *F. aquilonia* in Finland.

Our most striking result concerning the demographic history between the Finnish populations is the inference of a different pattern of gene flow between *F. polyctena* and *F. aquilonia*. There are two possible, non-mutually exclusive, causes for the bidirectional gene flow we now observe, one of which is direct and the other indirect. Direct introgression of alleles from *F. polyctena* into *F. aquilonia* could be facilitated by man-made close contact between individuals of these species. The forest management strategy practiced in Finland results in the formation of sharp boundaries between areas more suitable for *F. aquilonia* (forest interior with suitable temperature, shade and humidity) and areas where *F. polyctena* can thrive, as it can withstand increased exposure to sunlight (Punttila, 2020). The production of *F. aquilonia* sexual offspring has been described to be impaired both in deforested areas (Sorvari and Hakkarainen, 2007) and near the forest edge (Sorvari, 2013). *F. aquilonia* is commonly described in literature as a highly polygynous, highly polydomic, supercolonial species. As such, matings very often happen between individuals from the same nest, without any nuptial flight. On the assumption that deforestation and proximity to forest edge would reduce the number of in-nest sexuals, we could say that this could facilitate heterospecific matings due to lack of conspecific options, most likely with *F. polyctena* males mating with *F. aquilonia* females.

The second cause for the bidirectional gene flow is indirect. The *F. polyctena* x *F. aquilonia* hybrid populations extant in Southern Finland (Beresford *et al.*, 2017) could mediate gene flow from *F. polyctena* to *F. aquilonia* via backcrosses between hybrids and *F. aquilonia* individuals. This would lead to the introgression of *F. polyctena* genetic material into the *F. aquilonia* gene pool in Finland. We tested this scenario by considering a demographic model where the hybrid population continuously backcrosses with the parental species. Our results suggest that this is not the best model to explain the observed site frequency spectrum. The simple "Admixture" model, containing no backcrosses, was the best fit irrespective of the hybrid population considered. This could be simply because the data does not point towards the occurrence of backcrosses between hybrid and parental individuals, or because the parameter estimates of the simple "Admixture" model fit the data better, therefore increasing its likelihood. We may further investigate this by testing this model again with a pool of all hybrid samples or without allowing for any direct migration between the parental species. In any case, we would need to sample more pure Finnish representatives of both parental species, i.e. individuals that are not admixed with any other *Formica* species.

4.3. Is there evidence for gene flow from unsampled species into either *Formica polyctena* or *Formica aquilonia*?

Recently, evidence of gene flow and admixture with unsampled species has become more frequent (e.g., Kuhlwilm *et al.*, 2019). As it is known that both the species considered in this project, *F. polyctena* and *F. aquilonia*, may hybridize with closely related species of *rufa* group ants (Seifert and Goropashnaya, 2004; Seifert, Kulmuni and Pamilo, 2010), we considered the possibility that gene flow with another unsampled, more related species could be happening and not being accounted for in our models, possibly creating the signal of migration between *F. polyctena* and *F. aquilonia*. Given the available information, the most likely scenarios include at least one of their sister species sending migrants into either *F. polyctena* or *F. aquilonia*. In these scenarios, the unsampled species would be *F. rufa* for *F. polyctena*, and *F. lugubris*/*F. paralugubris* for *F. aquilonia*.

Our results suggest that these scenarios are not able to better explain the observed patterns of migration between *F. polyctena* and *F. aquilonia* than the models that include no unsampled species. However, the "(*F. polyctena*, Ghost), *F. aquilonia*" model revealed a consistent pattern of migration from the unsampled population, which would be *F. rufa* in this case, into *F. polyctena*. While this pattern does not constitute evidence of gene flow from an unsampled species into *F. polyctena*, it does warrant further investigation. We could explore this possibility by sampling *F. rufa* individuals and testing the same model using observed, sampled information from *F. rufa*.

4.4. How did the hybrid populations originate?

Evidence of admixture between *F. polyctena* and *F. aquilonia* in Southern Finland was first reported by Rosengren (1977), who identified *F. aquilonia* morphological traits in otherwise *F. polyctena*-like queens. Later, Sorvari (2006) sought to describe the same phenomenon in the worker caste, finding two separate morphological forms in *F. polyctena* workers. One of the forms presented a higher number of hairs than *F. polyctena* typically does, taking after the typical hairier *F. aquilonia* morphology. Sorvari (2006) was the first to postulate that *F. polyctena* and *F. aquilonia* may hybridise in Southern Finland. In Kulmuni, Seifert and Pamilo (2010), an established population of *F. polyctena* x *F. aquilonia* hybrids was morphologically and genetically described for the first time. This hybrid population was found to contain two separate genetic groups, corresponding to the Långholmen R and W subpopulations under study in this project. Over the years, more locations in Southern Finland have been found to be composed of *F. polyctena* x *F. aquilonia* hybrids (Beresford *et al.*, 2017). Thus far, hybridization between *F. polyctena* and *F. aquilonia* has been studied using allozyme, mitochondrial and microsatellite markers (Korczyńska *et al.*, 2010; Kulmuni, Seifert and Pamilo, 2010; Kulmuni and Pamilo, 2014; Beresford *et al.*, 2017). *F. polyctena* x *F. aquilonia* hybrid populations have only been studied as they were in the present or very recent time, as inference of their demographic history has never been attempted before. This project is the first to employ whole-genome data to not only study *F. polyctena* x *F. aquilonia* hybrid populations as they are in the present, but to also explore their origins in the past.

We found that admixture (i.e., hybridization) between *F. polyctena* and *F. aquilonia* is at the origin of our four sampled hybrid populations. The origins of all hybrid populations are estimated to be very recent and similar across populations. All hybrid populations received more of their genetic material from *F. polyctena* than from *F. aquilonia*, with *F. polyctena* contributing 58% to 68%. Our results suggest that the two Långholmen populations share a common origin, followed by several years of shared history before their ancestral population split into the two populations we find today. We cannot, however, state with certainty whether any of the other hybrid populations share common origins or not.

For the other hybrid populations, the expected likelihood values of the "Single Origin" and "Independent Origins" models are very similar, and, rather than supporting a single origin, the parameter estimates of the "Single Origin" model can also be consistent with two independent origins that happened at similar times and with similar contributions from the parental species. Indeed, it might be challenging to disentangle recent events that happened roughly simultaneously with SFS-based methods. Our estimates of genetic differentiation (F_{ST}), and our PCA and sNMF results also point to genetic differentiation between most hybrid populations (Table 3.2; Fig. 3.3-3.5), except for the two lineages in Långholmen, which is consistent with independent origins. In the future, we may employ methods to date admixture events based on LD-patterns once phased data becomes available, which might be more powerful to detect recent events than SFS-based methods (e.g., Sousa and Hey, 2013; Duranton *et al.*, 2018).

Compared to parental populations, the hybrid populations are quite genetically variable. The hybrid populations have generally higher genetic diversity than the populations of their parental species, which is not surprising in admixed individuals (e.g., Smith, Konings and Kornfield, 2003). All hybrid populations have negative F_{IS}, which we can partly attribute to selection or to recent hybridization (outcrossing). Kulmuni and Pamilo (2014) reported that in a *F. polyctena* x *F. aquilonia* hybrid population, due to differences in ploidy, genotype combinations that are selected against in males are favoured in females when they are heterozygous. This leads to an increase in the frequency of heterozygotes in these populations past what we would expect under HWE. The hybrid populations are fairly genetically different from each other, with the exception of the lineages in Långholmen, which show very limited differentiation between each other. However, while the hybrid individuals are clearly genetically intermediate between the parental species, they are more similar to each other than to individuals of their parental species. This is consistent with previous observations of *F. polyctena* x *F. aquilonia* hybrid individuals, which were found to be genetically more similar to each other than to pure individuals of their parental species (Korczyńska *et al.*, 2010). Interestingly, pairwise F_{ST} estimates and ancestry proportions reconstructed under K=2 indicate that the hybrid populations seem to be more similar to the *F. polyctena* populations, especially the one sampled in Finland, than to *F. aquilonia* populations. This supports our inference that *F. polyctena* contributed more genetic material to the hybrid populations than *F. aquilonia*. Analyses performed with different approaches, such as chromosome painting, also corroborate this observation (Nouhaud *et al*., in preparation). The hybrid populations are much more genetically similar to the populations of the parental species sampled in Finland than to those sampled outside Finland. Accordingly, this offers further support to the hypothesis that these populations result from admixture between *F. polyctena* and *F. aquilonia* in Southern Finland.

Hybridization in Formicidae ants is now known to be much more common than previously thought. For instance, recent studies have identified and described hybridization between the species *Tetramorium immigrans* and *T. caespitum* (Cordonnier *et al.*, 2019), and *Camponotus herculeanus* and *C. ligniperda* (Seifert, 2019). Importantly, hybridization between other *Formica* species has also been previously described, between both *rufa* and non-*rufa* group species. Seifert, Kulmuni and Pamilo (2010) reported frequent hybridization between *F. polyctena* and *F. rufa* in Central Europe, with hybrid individuals appearing to be genetically more similar to *F. polyctena* than *F. rufa*. Akin to our findings, hybrid individuals resulting from admixture between *F. selysi* and *F. cinerea* (Purcell *et al.*, 2016) are also predominantly genetically closer to one of the parental species, *F. selysi* in this case, than to the other. The available information in the literature, combined with the findings of this project, seems to suggest that hybridization between *Formica* species tends to happen asymmetrically, with one of the hybridizing species contributing more genetic material to the hybrid individuals than the others. Hybridization between other Hymenoptera species has been found to be both asymmetrical (e.g., Francisco *et al.*, 2014; Wallberg *et al.*, 2014) and non-asymmetrical (Anderson, Novak and Smith, 2008). Interestingly, theoretical work has predicted that biased introgression of mitochondrial genes via hybridization is

expected in haplodiploid organisms (Patten, Carioscia and Linnen, 2015). The genetic contributions of each of our parental species into the hybrids are asymmetrical in an approximately 60/40 ratio, this may be due to stochastic factors, such as backcrossing with the most abundant parental species in the area. It may be interesting to infer such parameters for other hybridizing haplodiploid species, such as those detailed in Nouhaud *et al.*, 2020).

References

Abbott, R. *et al.* (2013) 'Hybridization and speciation', *Journal of Evolutionary Biology*, 26(2), pp. 229–246. doi: 10.1111/j.1420-9101.2012.02599.x.

Anderson, K. E., Novak, S. J. and Smith, J. F. (2008) 'Populations composed entirely of hybrid colonies: Bidirectional hybridization and polyandry in harvester ants', *Biological Journal of the Linnean Society*, 95(2), pp. 320–336. doi: 10.1111/j.1095-8312.2008.01051.x.

Baack, E. J. and Rieseberg, L. H. (2007) 'A genomic view of introgression and hybrid speciation', *Current Opinion in Genetics and Development*, 17(6), pp. 513–518. doi: 10.1016/j.gde.2007.09.001.

Beichman, A. C., Huerta-Sanchez, E. and Lohmueller, K. E. (2018) 'Using genomic data to infer historic population dynamics of nonmodel organisms', *Annual Review of Ecology, Evolution, and Systematics*, 49, pp. 433–456. doi: 10.1146/annurev-ecolsys-110617-062431.

Beresford, J. *et al.* (2017) 'Widespread hybridization within mound-building wood ants in Southern Finland results in cytonuclear mismatches and potential for sex-specific hybrid breakdown', *Molecular Ecology*, 26(15), pp. 4013–4026. doi: 10.1111/mec.14183.

Blaimer, B. B. *et al.* (2015) 'Phylogenomic methods outperform traditional multi-locus approaches in resolving deep evolutionary history: A case study of formicine ants', *BMC Evolutionary Biology*. BMC Evolutionary Biology, 15(1), pp. 1–14. doi: 10.1186/s12862-015-0552-5.

Bolger, A. M., Lohse, M. and Usadel, B. (2014) 'Trimmomatic: A flexible trimmer for Illumina sequence data', *Bioinformatics*, 30(15), pp. 2114–2120. doi: 10.1093/bioinformatics/btu170.

Boomsma, J. J. *et al.* (2017) 'The Global Ant Genomics Alliance (GAGA)', *Myrmecological News*, 25(October), pp. 61–66.

Brelsford, A. *et al.* (2020) 'An Ancient and Eroded Social Supergene Is Widespread across Formica Ants', *Current Biology*, 30(2), pp. 304-311.e4. doi: 10.1016/j.cub.2019.11.032.

Burton, R. S., Pereira, R. J. and Barreto, F. S. (2013) 'Cytonuclear genomic interactions and hybrid breakdown', *Annual Review of Ecology, Evolution, and Systematics*, 44(November), pp. 281–302. doi: 10.1146/annurev-ecolsys-110512-135758.

Chan, K. O. *et al.* (2017) 'Species delimitation with gene flow: A methodological comparison and population genomics approach to elucidate cryptic species boundaries in Malaysian Torrent Frogs', *Molecular Ecology*, 26(20), pp. 5435–5450. doi: 10.1111/mec.14296.

Chapuisat, M. (1996) 'Characterization of microsatellite loci in Formica lugubris B and their variability in other ant species', *Molecular Ecology*, 5(4), pp. 599–601. doi: 10.1111/j.1365-294X.1996.tb00354.x.

Chapuisat, M., Bocherens, S. and Rosset, H. (2004) 'Variable queen number in ant colonies: No impact on queen turnover, inbreeding, and population genetic differentiation in the ant Formica selysi', *Evolution*, 58(5), pp. 1064–1072. doi: 10.1111/j.0014-3820.2004.tb00440.x.

Cordonnier, M. *et al.* (2019) 'From hybridization to introgression between two closely related sympatric ant species', *Journal of Zoological Systematics and Evolutionary Research*, 57(4), pp. 778–788. doi: 10.1111/jzs.12297.

Coyne, J. A. and Orr, A. H. (2004) 'Species: Reality and concepts', *Speciation*, p. 545.

Currat, M. *et al.* (2008) 'The hidden side of invasions: Massive introgression by local genes', *Evolution*,

62(8), pp. 1908–1920. doi: 10.1111/j.1558-5646.2008.00413.x.

Duranton, M. *et al.* (2018) 'The origin and remolding of genomic islands of differentiation in the European sea bass', *Nature Communications*. Springer US, 9(1), pp. 1–11. doi: 10.1038/s41467-018-04963-6.

Excoffier, L. *et al.* (2013) 'Robust Demographic Inference from Genomic and SNP Data', *PLoS Genetics*, 9(10). doi: 10.1371/journal.pgen.1003905.

Feldhaar, H., Foitzik, S. and Heinze, J. (2008) 'Review. Lifelong commitment to the wrong partner: Hybridization in ants', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1505), pp. 2891–2899. doi: 10.1098/rstb.2008.0022.

Filatov, D. A., Osborne, O. G. and Papadopulos, A. S. T. (2016) 'Demographic history of speciation in a Senecio altitudinal hybrid zone on Mt. Etna', *Molecular ecology*, 25(11), pp. 2467–2481. doi: 10.1111/mec.13618.

Francisco, F. O. *et al.* (2014) 'Hybridization and asymmetric introgression between Tetragonisca angustula and Tetragonisca fiebrigi', *Apidologie*, 45(1), pp. 1–9. doi: 10.1007/s13592-013-0224-7.

Frichot, E. *et al.* (2014) 'Fast and efficient estimation of individual ancestry coefficients', *Genetics*, 196(4), pp. 973–983. doi: 10.1534/genetics.113.160572.

Frichot, E. and François, O. (2015) 'APPLICATION LEA : An R package for landscape and ecological association studies', pp. 925–929. doi: 10.1111/2041-210X.12382.

Garnier-Géré, P. and Chikhi, L. (2013) 'Population Subdivision, Hardy-Weinberg Equilibrium and the Wahlund Effect', *eLS*, (November). doi: 10.1002/9780470015902.a0005446.pub3.

Garrison, E. and Marth, G. (2012) 'Haplotype-based variant detection from short-read sequencing', pp. 1–9.

Ghenu, A. H. *et al.* (2018) 'Conflict between heterozygote advantage and hybrid incompatibility in haplodiploids (and sex chromosomes)', *Molecular Ecology*, 27(19), pp. 3935–3949. doi: 10.1111/mec.14482.

Goodwin, S., McPherson, J. D. and McCombie, W. R. (2016) 'Coming of age: Ten years of next-generation sequencing technologies', *Nature Reviews Genetics*. Nature Publishing Group, 17(6), pp. 333–351. doi: 10.1038/nrg.2016.49.

Goropashnaya, A. V. *et al.* (2004) 'Limited phylogeographical structure across Eurasia in two red wood ant species Formica pratensis and F. lugubris (Hymenoptera, Formicidae)', *Molecular Ecology*, 13(7), pp. 1849–1858. doi: 10.1111/j.1365-294X.2004.02189.x.

Goropashnaya, A. V. *et al.* (2007) 'Phylogeography and population structure in the ant Formica exsecta (Hymenoptera, Formicidae) across Eurasia as reflected by mitochondrial DNA variation and microsatellites', *Annales Zoologici Fennici*, 44(6), pp. 462–474.

Goropashnaya, A. V. *et al.* (2012) 'Phylogenetic relationships of Palaearctic Formica species (hymenoptera, Formicidae) based on mitochondrial cytochrome b sequences', *PLoS ONE*, 7(7). doi: 10.1371/journal.pone.0041697.

Goropashnaya, A. V., Fedorov, V. B. and Pamilo, P. (2004) 'Recent speciation in the Formica rufa group ants (Hymenoptera, Formicidae): Inference from mitochondrial DNA phylogeny', *Molecular Phylogenetics and Evolution*, 32(1), pp. 198–206. doi: 10.1016/j.ympev.2003.11.016.

Goropashnaya, A. V., Seppä, P. and Pamilo, P. (2001) 'Social and genetic characteristics of geographically isolated populations in the ant Formica cinerea', *Molecular Ecology*, 10(12), pp. 2807–2818. doi: 10.1046/j.0962-1083.2001.01410.x.

Grimaldi, D. and Agosti, D. (2000) 'A formicine in New Jersey Cretaceous amber (Hymenoptera: Formicidae) and early evolution of the ants', *Proceedings of the National Academy of Sciences of the United States of America*, 97(25), pp. 13678–13683. doi: 10.1073/pnas.240452097.

Gyllenstrand, N., Gertsch, P. J. and Pamilo, P. (2002) 'Polymorphic microsatellite DNA markers in the ant Formica exsecta', *Molecular Ecology Notes*, 2(1), pp. 67–69. doi: 10.1046/j.1471-8286.2002.00152.x.

Hotaling, S. *et al.* (2018) 'Demographic modelling reveals a history of divergence with gene flow for a glacially tied stonefly in a changing post-Pleistocene landscape', *Journal of Biogeography*, 45(2), pp. 304–317. doi: 10.1111/jbi.13125.

Korczyńska, J. *et al.* (2010) 'Genetic polymorphism in "mixed" colonies of wood ants (Hymenoptera: Formicidae) in southern Finland and its possible origin', *European Journal of Entomology*, 107(2), pp. 157–167. doi: 10.14411/eje.2010.021.

Kuhlwilm, M. *et al.* (2019) 'Ancient admixture from an extinct ape lineage into bonobos', *Nature Ecology and Evolution*. Springer US, 3(6), pp. 957–965. doi: 10.1038/s41559-019-0881-7.

Kulmuni, J. and Pamilo, P. (2014) 'Introgression in hybrid ants is favored in females but selected against in males', *Proceedings of the National Academy of Sciences of the United States of America*, 111(35), pp. 12805–12810. doi: 10.1073/pnas.1323045111.

Kulmuni, J., Seifert, B. and Pamilo, P. (2010) 'Segregation distortion causes large-scale differences between male and female genomes in hybrid ants', 107(16). doi: 10.1073/pnas.0912409107.

De La Filia, A. G., Bain, S. A. and Ross, L. (2015) 'Haplodiploidy and the reproductive ecology of Arthropods', *Current Opinion in Insect Science*, 9, pp. 36–43. doi: 10.1016/j.cois.2015.04.018.

Li, H. and Durbin, R. (2010) 'Fast and accurate long-read alignment with Burrows-Wheeler transform', *Bioinformatics*, 26(5), pp. 589–595. doi: 10.1093/bioinformatics/btp698.

Liu, H. *et al.* (2017) 'Direct determination of the mutation rate in the bumblebee reveals evidence for weak recombination-associated mutation and an approximate rate constancy in insects', *Molecular Biology and Evolution*, 34(1), pp. 119–130. doi: 10.1093/molbev/msw226.

Mallet, J. (2007) 'Hybrid speciation', *Nature*, 446(7133), pp. 279–283. doi: 10.1038/nature05706.

Mallet, J. *et al.* (2009) 'Space, sympatry and speciation', *Journal of Evolutionary Biology*, 22(11), pp. 2332–2341. doi: 10.1111/j.1420-9101.2009.01816.x.

Nouhaud, P. *et al.* (2020) 'Understanding Admixture: Haplodiploidy to the Rescue', *Trends in Ecology & Evolution*, 35(1), pp. 34–42. doi: 10.1016/j.tree.2019.08.013.

Oswald, J. A. *et al.* (2017) 'Isolation with asymmetric gene flow during the nonsynchronous divergence of dry forest birds', *Molecular Ecology*, 26(5), pp. 1386–1400. doi: 10.1111/mec.14013.

Pamilo, P. (1982) 'Genetic population structure in polygynous formica ants', *Heredity*, 48(1), pp. 95–106. doi: 10.1038/hdy.1982.10.

Patten, M. M., Carioscia, S. A. and Linnen, C. R. (2015) 'Biased introgression of mitochondrial and

nuclear genes: A comparison of diploid and haplodiploid systems', *Molecular Ecology*, 24(20), pp. 5200–5210. doi: 10.1111/mec.13318.

Pfennig, K. S., Kelly, A. L. and Pierce, A. A. (2016) 'Hybridization as a facilitator of species range expansion', *Proceedings. Biological sciences*, 283(1839). doi: 10.1098/rspb.2016.1329.

Punttila, P. (2020) *Ant community structure in successional mosaics of boreal forests*.

Purcell, J. *et al.* (2016) 'Ants exhibit asymmetric hybridization in a mosaic hybrid zone', *Molecular Ecology*, 25(19), pp. 4866–4874. doi: 10.1111/mec.13799.

Rosengren, R. (1977) 'Foraging strategy of wood ants (Formica rufa group). I. Age polyethism and topographic traditions', *Acta Zoologica Fennica*, 149, pp. 1–33.

Ru, D. *et al.* (2018) 'Population genomic analysis reveals that homoploid hybrid speciation can be a lengthy process', *Molecular Ecology*, 27(23), pp. 4875–4887. doi: 10.1111/mec.14909.

Schwenk, K., Brede, N. and Streit, B. (2008) 'Introduction. Extent, processes and evolutionary impact of interspecific hybridization in animals', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1505), pp. 2805–2811. doi: 10.1098/rstb.2008.0055.

Seifert, B. (2019) 'Hybridization in the European carpenter ants Camponotus herculeanus and C. ligniperda (Hymenoptera: Formicidae)', *Insectes Sociaux*. Springer International Publishing, 66(3), pp. 365–374. doi: 10.1007/s00040-019-00693-0.

Seifert, B. and Goropashnaya, A. V. (2004) 'Ideal phenotypes and mismatching haplotypes - Errors of mtDNA treeing in ants (Hymenoptera: Formicidae) detected by standardized morphometry', *Organisms Diversity and Evolution*, 4(4), pp. 295–305. doi: 10.1016/j.ode.2004.04.005.

Seifert, B., Kulmuni, J. and Pamilo, P. (2010) 'Independent hybrid populations of Formica polyctena X rufa wood ants (Hymenoptera: Formicidae) abound under conditions of forest fragmentation', *Evolutionary Ecology*, 24(5), pp. 1219–1237. doi: 10.1007/s10682-010-9371-8.

Seppä, P. *et al.* (2012) 'Mosaic structure of native ant supercolonies', *Molecular Ecology*, 21(23), pp. 5880–5891. doi: 10.1111/mec.12070.

Smith, P. F., Konings, A. and Kornfield, I. (2003) 'Hybrid origin of a cichlid population in Lake Malawi: implications for genetic variation and species diversity', *Molecular Ecology*, 12(9), pp. 2497–2504. doi: 10.1046/j.1365-294X.2003.01905.x.

Sorvari, J. (2006) 'Two distinct morphs in the wood ant Formica polyctena in Finland: A result of hybridization?', *Entomologica Fennica*, 17(1), pp. 1–7. doi: 10.33338/ef.84281.

Sorvari, J. (2013) 'Proximity to the forest edge affects the production of sexual offspring and colony survival in the red wood ant Formica aquilonia in forest clear-cuts', *Scandinavian Journal of Forest Research*, 28(5), pp. 451–455. doi: 10.1080/02827581.2013.766258.

Sorvari, J. and Hakkarainen, H. (2007) 'Wood ants are wood ants: Deforestation causes population declines in the polydomous wood ant Formica aquilonia', *Ecological Entomology*, 32(6), pp. 707–711. doi: 10.1111/j.1365-2311.2007.00921.x.

Sousa, V. and Hey, J. (2013) 'Understanding the origin of species with genome-scale data: Modelling gene flow', *Nature Reviews Genetics*. Nature Publishing Group, 14(6), pp. 404–414. doi: 10.1038/nrg3446.

Steiner, F. M. *et al.* (2011) 'Mixed colonies and hybridisation of Messor harvester ant species (Hymenoptera: Formicidae)', *Organisms Diversity and Evolution*, 11(2), pp. 107–134. doi: 10.1007/s13127-011-0045-3.

Wakeley, J. (2004) 'Metapopulation models for historical inference', *Molecular Ecology*, 13(4), pp. 865–875. doi: 10.1111/j.1365-294X.2004.02086.x.

Wakeley, J. and Aliacar, N. (2001) 'Gene genealogies in a metrapop', *Genetics*, 159(1997), pp. 893–905.

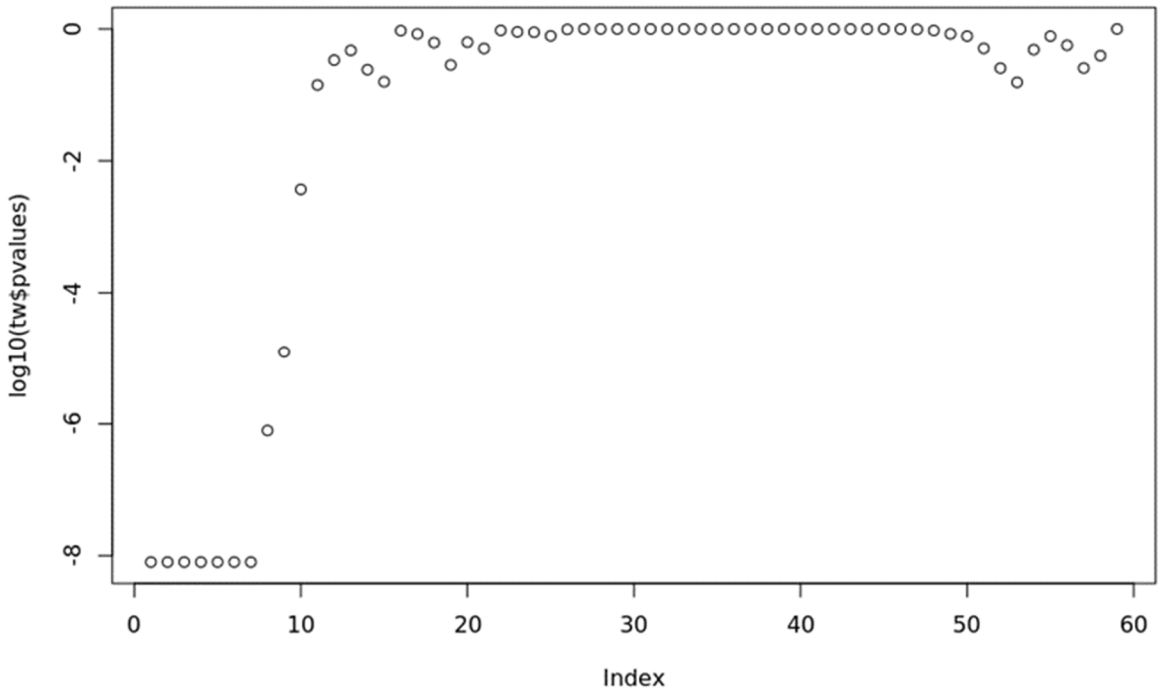
Wallberg, A. *et al.* (2014) 'A worldwide survey of genome sequence variation provides insight into the evolutionary history of the honeybee Apis mellifera', *Nature Genetics*. Nature Publishing Group, 46(10), pp. 1081–1088. doi: 10.1038/ng.3077.

Weir, B. S. and Cockerham, C. C. (1984) 'Estimating F-Statistics for the Analysis of Population Structure', *Evolution*, 38(6), p. 1358. doi: 10.2307/2408641.

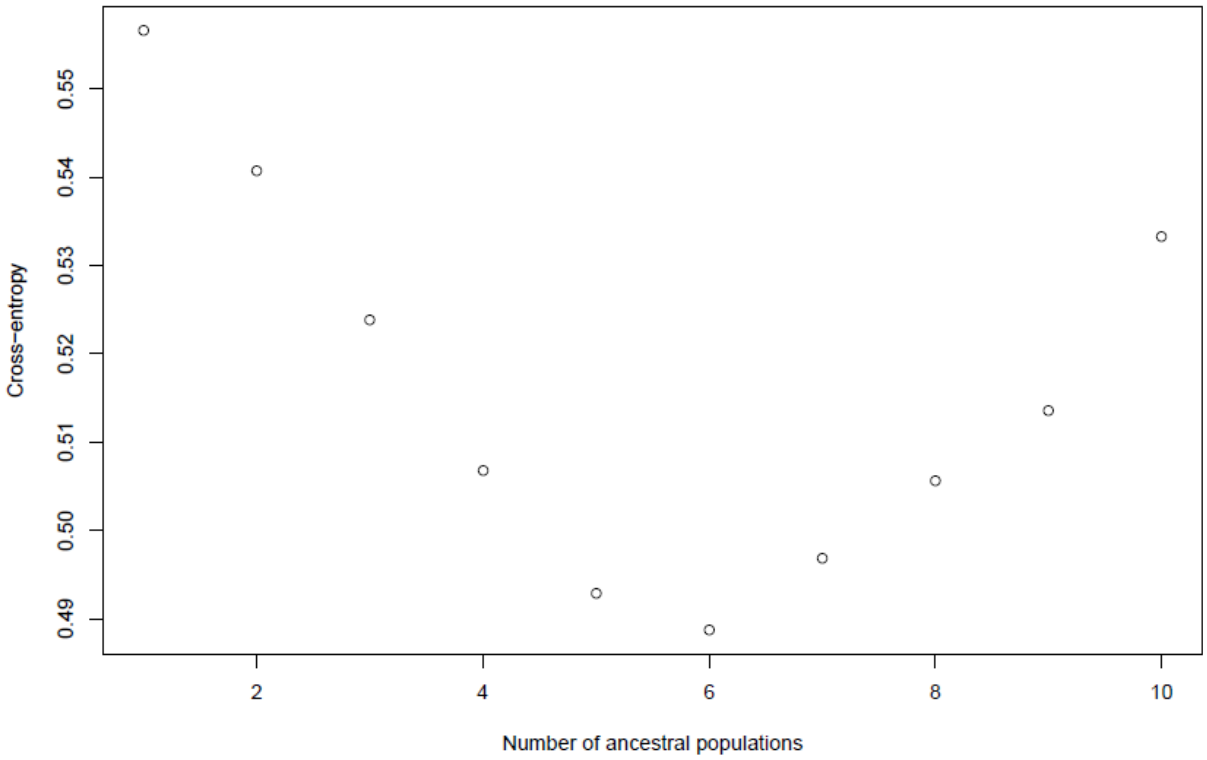
Welch, J. J. and Jiggins, C. D. (2014) 'Standing and flowing: The complex origins of adaptive variation', *Molecular Ecology*, 23(16), pp. 3935–3937. doi: 10.1111/mec.12859.

Zheng, X. *et al.* (2012) 'A high-performance computing toolset for relatedness and principal component analysis of SNP data', *Bioinformatics*, 28(24), pp. 3326–3328. doi: 10.1093/bioinformatics/bts606.

Supplementary Material



Supplementary Figure 1 - Tracy-Widom statistic applied to the Principal Components (PCs) assembled by the Principal Component Analysis. The first seven PCs are determined to be statistically significant.



Supplementary Figure 2 - Cross-entropy analysis for determination of the best number of ancestral clusters in the sNMF analysis. K=6 is determined to be the best number of ancestral clusters.

Supplementary Table 1 – Demographic parameters estimated by fastsimcoal2 in demographic model analyses. Unless bounded, the upper limit of the search range can be exceeded. Each model used only a subset of these parameters. Asterisks (*) mark parameters of models used to study the speciation history whose search ranges were altered in the "Sympatry" and "Migration after Isolation" when testing them with the Finnish comparison. The alternative minimum and maximum bounds are displayed in the appropriate columns. Double asterisks (**) mark parameters whose calculation changes between models.

				Search Range		
	Parameter	**Value Type**	**Distribution Type**	**Minimum**	**Maximum**	**Bounded?**
Speciation history models	N_ANC	Integer	Uniform	10	$2.0 * 10^6$	No
	N_ANC0*	Integer	Uniform	10; $2.0 * 10^5$	$2.0 * 10^6$; $4.0 * 10^5$	No
	N_ANC1*	Integer	Uniform	10; $3.0 * 10^5$	$2.0 * 10^6$; $5.0 * 10^5$	No
	N_POP0	Integer	Uniform	10	$2.0 * 10^6$	No
	N_POP1	Integer	Uniform	10	$2.0 * 10^6$	No
	TDIV*	Integer	Uniform	10; $2.0 * 10^5$	$5.0 * 10^5$; $4.0 * 10^5$	No
	REL_BOT	Float	Uniform	0	1	Yes
	REL_MIG	Float	Uniform	0	1	Yes
	T_BOT	Integer	Uniform	0	TDIV * REL_BOT	No
	TMIGSTART	Integer	Uniform	0	TDIV * REL_MIG	No
	TMIGSTOP	Integer	Uniform	0	TDIV * REL_MIG	No
	NM01	Float	Log-Uniform	$1.0 * 10^{-10}$	20	Yes
	NM10	Float	Log-Uniform	$1.0 *10^{-10}$	20	Yes
	MIG01**	Float	Log-Uniform	0	NM01/N_ANC0; NM01/N_POP0	No
	MIG10**	Float	Log-Uniform	0	NM10/N_ANC1; NM10/N_POP1	No
Unsampled species introgression models	N_ANC	Integer	Uniform	10	$2.0 * 10^6$	No
	N_ANC0	Integer	Uniform	10	$2.0 * 10^6$	No
	N_ANC1	Integer	Uniform	10	$2.0 * 10^6$	No
	N_POP0	Integer	Uniform	10	$2.0 * 10^6$	No
	N_POP1	Integer	Uniform	10	$2.0 * 10^6$	No
	N_GHOST	Integer	Uniform	10	$2.0 * 10^6$	No
	TDIV	Integer	Uniform	10	$5.0 * 10^5$	No
	REL_DIV	Float	Uniform	0	1	Yes
	TDIV_GHOST	Integer	Uniform	0	TDIV * REL_GHOST	No
	NM0GHOST	Float	Log-Uniform	$1.0 *10^{-10}$	20	Yes
	NM1GHOST	Float	Log-Uniform	$1.0 *10^{-10}$	20	Yes
	MIG0GHOST	Float	Log-Uniform	0	NM0GHOST/ N_POP0	No
	MIG1GHOST	Float	Log-Uniform	0	NM1GHOST/ N_POP1	No
	N_ANC_All	Integer	Uniform	$4.0 * 10^5$	$5.0 * 10^5$	No
	N_ANC0	Integer	Uniform	$2.0 * 10^5$	$3.0 * 10^5$	No
	N_ANCHYB	Integer	Uniform	10	$5.0 * 10^5$	No
	N_ANC2	Integer	Uniform	$3.0 * 10^5$	$5.0 * 10^5$	No
	N_POP0	Integer	Uniform	10	$3.0 * 10^5$	No
	N_HYB	Integer	Uniform	10	$3.0 * 10^5$	No
	N_POP2	Integer	Uniform	10	$3.0 * 10^5$	No
	N_POP0_REC	Integer	Uniform	10	$3.0 * 10^5$	No
	N_HYB_REC	Integer	Uniform	10	$3.0 * 10^5$	No
	N_POP2_REC	Integer	Uniform	10	$3.0 * 10^5$	No
	TDIV	Integer	Uniform	$2.0 * 10^5$	$3.0 * 10^5$	No
	REL_MIG	Float	Uniform	0	1	Yes

Secondary Contact and Admixture models	REL_DIV	Float	Uniform	0	1	Yes
	REL_BOT	Float	Uniform	0	1	Yes
	REL_ADM	Float	Uniform	0	1	Yes
	TMIGSTOP	Integer	Uniform	0	TDIV * REL_MIG	No
	TDIV01	Integer	Uniform	0	TDIV * REL_DIV	No
	TDIV12	Integer	Uniform	0	TDIV*REL_DIV	No
	TBOT**	Integer	Uniform	0	TMIGSTOP*REL_BOT; TADMS*REL_BOT	No
	TADMS	Integer	Uniform	0	TDIV*REL_ADM	No
	TADME	Integer	Uniform	0	TADMS+1	No
	NM01	Float	Log-Uniform	$1.0 * 10^{-10}$	20	Yes
	NM10	Float	Log-Uniform	$1.0 * 10^{-10}$	20	Yes
	NM12	Float	Log-Uniform	$1.0 * 10^{-10}$	20	Yes
	NM21	Float	Log-Uniform	$1.0 * 10^{-10}$	20	Yes
	NM02	Float	Log-Uniform	$1.0 * 10^{-10}$	20	Yes
	NM20	Float	Log-Uniform	$1.0 * 10^{-10}$	20	Yes
	NM02_ANC	Float	Log-Uniform	$1.0 * 10^{-10}$	20	Yes
	ALFA	Float	Uniform	0	1	Yes
	MIG01**	Float	Log-Uniform	0	NM01/N_POP0(_REC)	No
	MIG10**	Float	Log-Uniform	0	NM10/N_HYB(_REC)	No
	MIG12**	Float	Log-Uniform	0	NM12/N_HYB(_REC)	No
	MIG21**	Float	Log-Uniform	0	NM21/N_POP2(_REC)	No
	MIG02**	Float	Log-Uniform	0	NM02/N_POP0(_REC)	No
	MIG20**	Float	Log-Uniform	0	NM20/N_POP2(_REC)	No
	MIG02_ANC	Float	Log-Uniform	0	NM02_ANC/N_ANC0	No
Single Origin and Independent Origins	N_ANC_All	Integer	Uniform	$4.0 * 10^5$	$5.0 * 10^5$	No
	N_ANC0	Integer	Uniform	$2.0 * 10^5$	$3.0 * 10^5$	No
	N_ANC3	Integer	Uniform	$3.0 * 10^5$	$5.0 * 10^5$	No
	N_ANCHYB	Integer	Uniform	10	$2.0 * 10^3$	No
	N_POP0	Integer	Uniform	10	$3.0 * 10^5$	No
	N_POP3	Integer	Uniform	10	$3.0 * 10^5$	No
	N_POP0_REC	Integer	Uniform	10	$5.0 * 10^4$	No
	N_HYB1_REC	Integer	Uniform	10	$2.0 * 10^3$	No
	N_HYB2_REC	Integer	Uniform	10	$2.0 * 10^3$	No
	N_POP3_REC	Integer	Uniform	10	$5.0 * 10^4$	No
	TDIV	Integer	Uniform	$2.0 * 10^5$	$3.0 * 10^5$	No
	TADMS	Integer	Uniform	0	50	No
	TADMS_HYB1	Integer	Uniform	0	50	No
	TADMS_HYB2	Integer	Uniform	0	50	No
	REL_BOT	Float	Uniform	0	1	Yes
	REL_DIV	Float	Uniform	0	1	Yes
	TBOT	Integer	Uniform	0	TDIV*REL_BOT	No
	TADME	Integer	Uniform	0	TADMS+1	No
	TADME_HYB1	Integer	Uniform	0	TADMS_HYB1+1	No
	TADME_HYB2	Integer	Uniform	0	TADMS_HYB2+1	No
	TDIV_HYB	Integer	Uniform	0	TADMS*REL_DIV	No
	ALFA	Float	Uniform	0	1	Yes
	BETA	Float	Uniform	0	1	Yes
	NM02_ANC	Float	Log-Uniform	0	NM02_ANC/N_ANC0	No

Supplementary Table 2 - Maximum likelihood parameter estimates for all models concerning the speciation history between *Formica polyctena* and *Formica aquilonia*, tested with the dataset where the *F. polyctena* population in West Switzerland is the first population, and the *F. aquilonia* population in Scotland is the second. All effective sizes (Ne) are given in number of haploids. Times are given in number of generations. Migration rates are scaled according to population effective sizes (2Nm). Maximum-likelihood estimates for parameters are taken from the run reaching the highest composite likelihood of the 100 runs performed. Likelihoods are given in logarithmic scale. Maximum observed likelihood for this dataset is -2,037,909.731. ΔLikelihood is calculated by subtracting the expected likelihood from the maximum observed likelihood.

Parameter	Allopatry	Sympatry	Isolation after Migration	Migration after Isolation
Ancestral N_e	509,350	419,736	502,389	454,087
***F. polyctena* ancestral N_e**	1,454,254	284,346	387,423	392,188
***F. aquilonia* ancestral N_e**	174,024	465,565	364,037	507,140
***F. polyctena* N_e**	83,057	57,374	78,212	86,835
***F. aquilonia* N_e**	146,632	76,431	106,691	99,898
Time of divergence	110,374	297,231	220,636	188,521
Time of size change	47,069	26,342	-	-
Time of isolation	-	-	40,820	-
Time of contact	-	-	-	43,314
2*Nm* (*F. aquilonia* to *F. polyctena*)	-	0.9860377	2.6208967	0.2509777
2*Nm* (*F. polyctena* to *F. aquilonia*)	-	1.37E-05	2.65E-04	1.25E-08
Expected Likelihood	-2,038,390.527	-2,038,351.735	-2,038,351.735	-2,038,304.638
ΔLikelihood	480.796	442.004	442.004	394.907

Supplementary Table 3 - Maximum likelihood parameter estimates for all models concerning the speciation history between *Formica polyctena* and *Formica aquilonia*, tested with the dataset where the *F. polyctena* population in East Switzerland is the first population, and the *F. aquilonia* population in Scotland is the second. All effective sizes (Ne) are given in number of haploids. Times are given in number of generations. Migration rates are scaled according to population effective sizes (2Nm). Maximum-likelihood estimates for parameters are taken from the run reaching the highest composite likelihood of the 100 runs performed. Likelihoods are given in logarithmic scale. Maximum observed likelihood for this dataset is -2,163,676.544. ΔLikelihood is calculated by subtracting the expected likelihood from the maximum observed likelihood.

Parameter	**Allopatry**	**Sympatry**	**Isolation after Migration**	**Migration after Isolation**
Ancestral N_e	501,519	475,840	491,856	491,282
***F. polyctena* ancestral N_e**	665,985	299,620	139,822	1,549,097
***F. aquilonia* ancestral N_e**	236,637	366,316	339,032	310,438
***F. polyctena* N_e**	8,548	29,080	38,378	18,720
***F. aquilonia* N_e**	38,767	51,997	29,147	53,680
Time of divergence	135,229	207,033	220,545	154,813
Time of size change	4,688	12,916	-	-
Time of isolation	-	-	5,696	-
Time of contact	-	-	-	12,513
2*Nm* (*F. aquilonia* to *F. polyctena*)	-	0.5668583	0.3935858	0.0280896
2*Nm* (*F. polyctena* to *F. aquilonia*)	-	4.46E-04	1.10E-03	9.88E-06
Expected Likelihood	-2,164,109.333	-2,164,022.947	-2,164,046.421	-2,164,087.426
ΔLikelihood	432.789	346.403	369.88	410.882

Supplementary Table 4 - Maximum likelihood parameter estimates for all models concerning the speciation history between *Formica polyctena* and *Formica aquilonia*, tested with the dataset where the *F. polyctena* population in West Switzerland is the first population, and the *F. aquilonia* population in Switzerland is the second. All effective sizes (Ne) are given in number of haploids. Times are given in number of generations. Migration rates are scaled according to population effective sizes (2Nm). Maximum-likelihood estimates for parameters are taken from the run reaching the highest composite likelihood of the 100 runs performed. Likelihoods are given in logarithmic scale. Maximum observed likelihood for this dataset is -1,993,626.486. ΔLikelihood is calculated by subtracting the expected likelihood from the maximum observed likelihood.

Parameter	Allopatry	Sympatry	Isolation after Migration	Migration after Isolation
Ancestral N_e	487,733	477,516	485,433	511,431
***F. polyctena* ancestral N_e**	1,237,745	359,341	247,349	213,329
***F. aquilonia* ancestral N_e**	320,356	466,128	391,948	213,709
***F. polyctena* N_e**	42,459	74,192	65,244	86,025
***F. aquilonia* N_e**	62,808	70,679	75,344	76,323
Time of divergence	136,971	174,591	285,395	118,533
Time of size change	21,622	31,393	-	-
Time of isolation	-	-	27,083	-
Time of contact	-	-	-	20,352
2*Nm* (*F. aquilonia* to *F. polyctena*)	-	0.5433142	1.401815	0.1133943
2*Nm* (*F. polyctena* to *F. aquilonia*)	-	9.17E-03	1.30E-04	2.24E-06
Expected Likelihood	-1,994,231.487	-1,994,187.631	-1,994,132.769	-1,994,172.300
ΔLikelihood	605.001	561.145	506.283	545.814

Supplementary Table 5 - Maximum likelihood parameter estimates for all models concerning the speciation history between *Formica polyctena* and *Formica aquilonia*, tested with the dataset where the both populations were sampled in Finland. All effective sizes (Ne) are given in number of haploids. Times are given in number of generations. Migration rates are scaled according to population effective sizes (2Nm). Maximum-likelihood estimates for parameters are taken from the run reaching the highest composite likelihood of the 100 runs performed. Likelihoods are given in logarithmic scale. Maximum observed likelihood for this dataset is -1,877,036.056. ΔLikelihood is calculated by subtracting the expected likelihood from the maximum observed likelihood.

Parameter	Allopatry	Sympatry	Isolation after Migration	Migration after Isolation
Ancestral N_e	452,531	421,803	454,136	401,707
***F. polyctena* ancestral N_e**	1,562,309	205,477	248,080	264,026
***F. aquilonia* ancestral N_e**	70,034	312,051	189,624	435,786
***F. polyctena* N_e**	165,752	268,225	741,957	248,327
***F. aquilonia* N_e**	1,324,820	138,939	209,965	167,201
Time of divergence	55,771	224,481	115,730	204,610
Time of size change	24,642	39,277	-	-
Time of isolation	-	-	1,789	-
Time of contact	-	-	-	81,474
2*Nm* (*F. aquilonia* to *F. polyctena*)	-	1.2670534	1.2964164	1.7686041
2*Nm* (*F. polyctena* to *F. aquilonia*)	-	0.2044848	3.31E-03	0.1379156
Expected Likelihood	-1,877,356.775	-1,877,155.931	-1,877,194.356	-1,877,161.719
ΔLikelihood	320.719	119.875	158.300	125.663

Supplementary Table 6 - Maximum likelihood parameter estimates for the "(*Formica polyctena*, Ghost), *Formica aquilonia*" model tested with all population comparisons. All effective sizes (Ne) are given in number of haploids. Times are given in number of generations. Migration rates are scaled according to population effective sizes (2Nm). Maximum-likelihood estimates for parameters are taken from the run reaching the highest composite likelihood of the 100 runs performed. Likelihoods are given in logarithmic scale. ΔLikelihood is calculated by subtracting the expected likelihood from the maximum observed likelihood.

Parameter	West Switzerland *F. polyctena* + Scotland *F. aquilonia*	West Switzerland *F. polyctena* + Switzerland *F. aquilonia*	East Switzerland *F. polyctena* + Scotland *F. aquilonia*	Finland *F. polyctena* + Finland *F. aquilonia*
Ancestral N_e	483,222	506,147	495,331	445,090
***F. polyctena* ancestral N_e**	1,799,354	799,601	1,874,157	1,613,901
***F. aquilonia* ancestral N_e**	220,734	194,478	289,007	85,428
***F. polyctena* N_e**	9,271	9,245	22,638	40,944
***F. aquilonia* N_e**	50,335	110,562	72,189	219,833
Ghost population N_e	1,123,242	325,187	1,559,278	1,802,127
Time of divergence	130,207	118,229	146,102	64,003
Time of divergence (Ghost/*F. polyctena*)	5,902	27,425	15,015	40,473
2Nm (Ghost to *F. polyctena*)	0.2793092	0.9060956	0.000312282	3.1094804
Max. Observed Likelihood	-2,037,909.731	-1,993,626.486	-2,163,676.544	-1,877,036.056
Expected Likelihood	-2,038,344.718	-1,994,224.792	-2,164,118.547	-1,877,320.70
ΔLikelihood	394.907	598.306	442.003	284.642

Supplementary Table 7 - Maximum likelihood parameter estimates for the "(*Formica aquilonia*, Ghost), *Formica polyctena*" model tested with all population comparisons. All effective sizes (Ne) are given in number of haploids. Times are given in number of generations. Migration rates are scaled according to population effective sizes (2Nm). Maximum-likelihood estimates for parameters are taken from the run reaching the highest composite likelihood of the 100 runs performed. Likelihoods are given in logarithmic scale. ΔLikelihood is calculated by subtracting the expected likelihood from the maximum observed likelihood.

Parameter	West Switzerland F. polyctena + Scotland F. aquilonia	West Switzerland *F. polyctena* + Switzerland *F. aquilonia*	East Switzerland *F. polyctena* + Scotland *F. aquilonia*	Finland *F. polyctena* + Finland *F. aquilonia*
Ancestral N_e	504,126	501,774	491,693	456,362
***F. polyctena* ancestral N_e**	1,942,980	352,775	997,620	1,323,820
***F. aquilonia* ancestral N_e**	195,253	217,287	299,329	76,533
***F. polyctena* N_e**	90,008	59,116	11,122	184,811
***F. aquilonia* N_e**	141,283	80,754	36,280	303,743
Ghost population N_e	1,748,768	1,316,349	1,496,129	427,941
Time of divergence	112,989	120,368	152,108	56,728
Time of divergence (Ghost/*F. aquilonia*)	51,687	20,437	6,647	29,225
2Nm (Ghost to *F. aquilonia*)	1.91E-08	0.000310889	0.0215419	2.41E-05
Max. Observed Likelihood	-2,037,909.731	-1,993,626.486	-2,163,676.544	-1,877,036.056
Expected Likelihood	-2,038,394.770	-1,994,228.065	-2,164,124.248	-1,877,354.99
ΔLikelihood	485.039	601.579	447.704	318.930

Supplementary Table 8 - Maximum likelihood parameter estimates for the Secondary Contact models concerning the origin of the hybrid populations, tested with the dataset containing Bunkkeri and the Finnish *Formica polyctena* and *Formica aquilonia* parental populations. All effective sizes (Ne) are given in number of haploids. Times are given in number of generations. Migration rates are scaled according to population effective sizes (2Nm). Maximum-likelihood estimates for parameters are taken from the run reaching the highest composite likelihood of the 100 runs performed. Likelihoods are given in logarithmic scale. Maximum observed likelihood for this dataset is -1,821,814.189. ΔLikelihood is calculated by subtracting the expected likelihood from the maximum observed likelihood.

Parameter	Trifurcation	(*F. polyctena*, Hybrid), *F. aquilonia*	(*F. aquilonia*, Hybrid), *F. polyctena*
Ancestral N_e	404,359	407,430	406,698
***F. polyctena* ancestral N_e**	213,851	203,991	209,753
Hybrid ancestral N_e	394,642	-	-
***F. aquilonia* ancestral N_e**	304,766	304,103	308,406
"Older" *F. polyctena* N_e	-	96,556	50,490
"Older" hybrid N_e	-	210,037	158,705
"Older" *F. aquilonia* Ne	-	68,802	118,475
Recent *F. polyctena* Ne	59,117	27,996	22,045
Recent Hybrid N_e	20,187	17,962	8,823
Recent *F. aquilonia* N_e	88,710	82,700	9,464
Time of divergence	259,246	204,399	202,275
Time of divergence (Hybrid from *F. polyctena*)	-	22,394	-
Time of divergence (Hybrid from *F. aquilonia*)	-	-	17,583
Time of contact	24,172	20,708	15,588
Time of size change	-	3,105	1,298
2Nm (Hybrid to *F. polyctena*)	0.0034056	-	4.12E-07
2Nm (*F. polyctena* to Hybrid)	0.9529539	-	5.25502
2Nm (*F. aquilonia* to Hybrid)	1.0634894	5.8546539	-
2Nm (Hybrid to *F. aquilonia*)	0.0017806	3.10E-08	-
2Nm (*F. aquilonia* to *F. polyctena*)	0.0014497	0.0027863	0.026998
2Nm (*F. polyctena* to *F. aquilonia*)	0.0024855	0.0658583	4.76E-09
Ancestral 2Nm (*F. aquilonia* to *F. polyctena*)	1.3221118	0.8385864	1.0483054
LogLikelihood	-1,823,911.477	-1,823,310.938	-1,823,193.952
ΔLikelihood	2,097.288	1,496.749	1,379.763

Supplementary Table 9 - Maximum likelihood parameter estimates for the Admixture models concerning the origin of the hybrid populations, tested with the dataset containing Bunkkeri and the Finnish *Formica polyctena* and *Formica aquilonia* parental populations. All effective sizes (Ne) are given in number of haploids. Times are given in number of generations. Migration rates are scaled according to population effective sizes (2Nm). Maximum-likelihood estimates for parameters are taken from the run reaching the highest composite likelihood of the 100 runs performed. Likelihoods are given in logarithmic scale. Maximum observed likelihood for this dataset is -1,821,814.189. ΔLikelihood is calculated by subtracting the expected likelihood from the maximum observed likelihood.

Parameter	Admixture	Admixture with continuous migration
Ancestral N_e	404,615	405,991
***F. polyctena* ancestral N_e**	236,539	270,578
***F. aquilonia* ancestral N_e**	349,028	481,289
"Older" *F. polyctena* N_e	237,798	241,900
"Older" *F. aquilonia* Ne	242,562	245,992
Recent *F. polyctena* Ne	59	113
Recent Hybrid N_e	96	193
Recent *F. aquilonia* N_e	177	284
Time of divergence	204,687	203,593
Time of admixture	19	36
Time of size change	196,987	200,721
Genetic input from *F. polyctena*	0.5663739	0.5573837
Genetic input from *F. aquilonia*	0.4336261	0.4426163
2Nm (Hybrid to *F. polyctena*)	-	9.59E-06
2Nm (*F. polyctena* to Hybrid)	-	5.15E-07
2Nm (*F. aquilonia* to Hybrid)	-	3.74E-08
2Nm (Hybrid to *F. aquilonia*)	-	5.79E-08
Ancestral 2Nm (*F. aquilonia* to *F. polyctena*)	0.849027	0.9486129
LogLikelihood	-1,822,728.332	-1,822,730.080
ΔLikelihood	914.143	915.891

Supplementary Table 10 - Maximum likelihood parameter estimates for the Secondary Contact models concerning the origin of the hybrid populations, tested with the dataset containing Pikkala and the Finnish *Formica polyctena* and *Formica aquilonia* parental populations. All effective sizes (Ne) are given in number of haploids. Times are given in number of generations. Migration rates are scaled according to population effective sizes (2Nm). Maximum-likelihood estimates for parameters are taken from the run reaching the highest composite likelihood of the 100 runs performed. Likelihoods are given in logarithmic scale. Maximum observed likelihood for this dataset is -1,377,702.973. ΔLikelihood is calculated by subtracting the expected likelihood from the maximum observed likelihood.

Parameter	Trifurcation	(*F. polyctena*, Hybrid), *F. aquilonia*	(*F. aquilonia*, Hybrid), *F. polyctena*
Ancestral N_e	422,019	407,868	406,798
***F. polyctena* ancestral N_e**	225,707	212,502	233,524
Hybrid ancestral N_e	414,707	-	-
***F. aquilonia* ancestral N_e**	305,722	307,153	355,954
"Older" *F. polyctena* N_e	-	55,038	32,175
"Older" hybrid N_e	-	101,532	146,057
"Older" *F. aquilonia* Ne	-	69,287	43,639
Recent *F. polyctena* Ne	42,210	189,730	110,760
Recent Hybrid N_e	11,406	3,420	4,475
Recent *F. aquilonia* N_e	69,419	197,472	146,797
Time of divergence	268,899	237,289	272,773
Time of divergence (Hybrid from *F. polyctena*)	-	19,129	-
Time of divergence (Hybrid from *F. aquilonia*)	-	-	14,470
Time of contact	17,464	18,350	10,443
Time of size change	-	647	1,073
2Nm (Hybrid to *F. polyctena*)	3.64E-07	-	1.47E-07
2Nm (*F. polyctena* to Hybrid)	0.6424457	-	5.8603323
2Nm (*F. aquilonia* to Hybrid)	0.7183277	2.9147546	-
2Nm (Hybrid to *F. aquilonia*)	1.13E-05	3.66E-07	-
2Nm (*F. aquilonia* to *F. polyctena*)	5.74E-06	2.01E-06	1.98E-04
2Nm (*F. polyctena* to *F. aquilonia*)	0.0045192	0.0112385	2.80E-03
Ancestral 2Nm (*F. aquilonia* to *F. polyctena*)	1.366713	1.050823	1.4402665
LogLikelihood	-1,379,960.87	-1,378,640.59	-1,378,994.04
ΔLikelihood	2,257.901	937.613	1,291.063

Supplementary Table 11 - Maximum likelihood parameter estimates for the Admixture models concerning the origin of the hybrid populations, tested with the dataset containing Pikkala and the Finnish *Formica polyctena* and *Formica aquilonia* parental populations. All effective sizes (Ne) are given in number of haploids. Times are given in number of generations. Migration rates are scaled according to population effective sizes (2Nm). Maximum-likelihood estimates for parameters are taken from the run reaching the highest composite likelihood of the 100 runs performed. Likelihoods are given in logarithmic scale. Maximum observed likelihood for this dataset is -1,377,702.973. ΔLikelihood is calculated by subtracting the expected likelihood from the maximum observed likelihood.

Parameter	Admixture	Admixture with continuous migration
Ancestral N_e	406,067	406,945
***F. polyctena* ancestral N_e**	223,299	273,118
***F. aquilonia* ancestral N_e**	355,254	347,083
"Older" *F. polyctena* N_e	291,274	291,695
"Older" *F. aquilonia* Ne	218,427	223,263
Recent *F. polyctena* Ne	103	129
Recent Hybrid N_e	134	153
Recent *F. aquilonia* N_e	337	357
Time of divergence	210,810	207,887
Time of admixture	37	43
Time of size change	113,472	137,129
Genetic input from *F. polyctena*	0.5888864	0.6155127
Genetic input from *F. aquilonia*	0.4111136	0.3844873
2Nm (Hybrid to *F. polyctena*)	-	2.92E-06
2Nm (*F. polyctena* to Hybrid)	-	5.15E-09
2Nm (*F. aquilonia* to Hybrid)	-	1.72E-04
2Nm (Hybrid to *F. aquilonia*)	-	0.3970062
Ancestral 2Nm (*F. aquilonia* to *F. polyctena*)	0.801646	0.8374673
LogLikelihood	-1,378,464.891	-1,378,474.028
ΔLikelihood	761.918	771.055

Supplementary Table 12 - Maximum likelihood parameter estimates for the Secondary Contact models concerning the origin of the hybrid populations, tested with the dataset containing Långholmen W and the Finnish *Formica polyctena* and *Formica aquilonia* parental populations. All effective sizes (Ne) are given in number of haploids. Times are given in number of generations. Migration rates are scaled according to population effective sizes (2Nm). Maximum-likelihood estimates for parameters are taken from the run reaching the highest composite likelihood of the 100 runs performed. Likelihoods are given in logarithmic scale. Maximum observed likelihood for this dataset is -1,144,658.203. ΔLikelihood is calculated by subtracting the expected likelihood from the maximum observed likelihood.

Parameter	**Trifurcation**	**(*F. polyctena*, Hybrid), *F. aquilonia***	**(*F. aquilonia*, Hybrid), *F. polyctena***
Ancestral N_e	413,444	404,415	405,879
***F. polyctena* ancestral N_e**	255,747	215,583	222,377
Hybrid ancestral N_e	401,225	-	-
***F. aquilonia* ancestral N_e**	305,592	305,473	314,629
"Older" *F. polyctena* N_e	-	64,194	55,900
"Older" hybrid N_e	-	177,575	131,325
"Older" *F. aquilonia* Ne	-	81,193	84,701
Recent *F. polyctena* Ne	38,971	91,607	84,617
Recent Hybrid N_e	10,158	8,323	3,854
Recent *F. aquilonia* N_e	77,805	251,424	218,405
Time of divergence	278,755	231,376	206,806
Time of divergence (Hybrid from *F. polyctena*)	-	21,738	-
Time of divergence (Hybrid from *F. aquilonia*)	-	-	24,726
Time of contact	16,288	20,780	23,305
Time of size change	-	2,266	851
2Nm (Hybrid to *F. polyctena*)	1.51E-05	-	0.3017295
2Nm (*F. polyctena* to Hybrid)	0.6570334	-	2.8048582
2Nm (*F. aquilonia* to Hybrid)	0.6774944	4.307785	-
2Nm (Hybrid to *F. aquilonia*)	3.9960E-05	1.27E-04	-
2Nm (*F. aquilonia* to *F. polyctena*)	1.25E-05	0.0016129	6.75E-04
2Nm (*F. polyctena* to *F. aquilonia*)	7.85E-07	4.29E-09	6.35E-08
Ancestral 2Nm (*F. aquilonia* to *F. polyctena*)	1.3760906	9.43E-01	7.69E-01
LogLikelihood	-1,149,183.231	-1,147,521.50	-1,147,467.010
ΔLikelihood	4,525.028	2,863.301	2,808.807

Supplementary Table 13 - Maximum likelihood parameter estimates for the Admixture models concerning the origin of the hybrid populations, tested with the dataset containing Långholmen W and the Finnish *Formica polyctena* and *Formica aquilonia* parental populations. All effective sizes (Ne) are given in number of haploids. Times are given in number of generations. Migration rates are scaled according to population effective sizes (2Nm). Maximum-likelihood estimates for parameters are taken from the run reaching the highest composite likelihood of the 100 runs performed. Likelihoods are given in logarithmic scale. Maximum observed likelihood for this dataset is -1,144,658.203. ΔLikelihood is calculated by subtracting the expected likelihood from the maximum observed likelihood.

Parameter	Admixture	Admixture with continuous migration
Ancestral N_e	408,106	406,205
***F. polyctena* ancestral N_e**	267,746	230,663
***F. aquilonia* ancestral N_e**	358,894	323,846
"Older" *F. polyctena* N_e	223,182	256,805
"Older" *F. aquilonia* Ne	256,382	224,806
Recent *F. polyctena* Ne	117	175
Recent Hybrid N_e	116	159
Recent *F. aquilonia* N_e	206	340
Time of divergence	233,179	214,228
Time of admixture	36	50
Time of size change	92,010	31,785
Genetic input from *F. polyctena*	0.5496706	0.5293458
Genetic input from *F. aquilonia*	0.4503294	0.4706542
2Nm (Hybrid to *F. polyctena*)	-	5.42E-06
2Nm (*F. polyctena* to Hybrid)	-	2.52E-06
2Nm (*F. aquilonia* to Hybrid)	-	4.59E-04
2Nm (Hybrid to *F. aquilonia*)	-	2.34E-07
Ancestral 2Nm (*F. aquilonia* to *F. polyctena*)	1.046214	0.7119682
LogLikelihood	-1,147,133.678	-1,147,134.384
ΔLikelihood	2,475.475	2,476.181

Supplementary Table 14 - Maximum likelihood parameter estimates for the Secondary Contact models concerning the origin of the hybrid populations, tested with the dataset containing Långholmen R and the Finnish *Formica polyctena* and *Formica aquilonia* parental populations. All effective sizes (Ne) are given in number of haploids. Times are given in number of generations. Migration rates are scaled according to population effective sizes (2Nm). Maximum-likelihood estimates for parameters are taken from the run reaching the highest composite likelihood of the 100 runs performed. Likelihoods are given in logarithmic scale. Maximum observed likelihood for this dataset is -883,471.568. ΔLikelihood is calculated by subtracting the expected likelihood from the maximum observed likelihood.

Parameter	Trifurcation	(*F. polyctena*, Hybrid), *F. aquilonia*	(*F. aquilonia*, Hybrid), *F. polyctena*
Ancestral N_e	406,159	411,696	406,395
***F. polyctena* ancestral N_e**	249,110	221,617	253,880
Hybrid ancestral N_e	359,959	-	-
***F. aquilonia* ancestral N_e**	311,983	314,804	360,014
"Older" *F. polyctena* N_e	-	95,334	41,184
"Older" hybrid N_e	-	208,042	104,447
"Older" *F. aquilonia* Ne	-	113,978	50,568
Recent *F. polyctena* Ne	24,492	152,241	152,780
Recent Hybrid N_e	5,975	2,133	564
Recent *F. aquilonia* N_e	36,689	185,953	188,210
Time of divergence	263,104	203,648	208,595
Time of divergence (Hybrid from *F. polyctena*)	-	43,288	-
Time of divergence (Hybrid from *F. aquilonia*)	-	-	15,198
Time of contact	9,096	35,385	12,813
Time of size change	-	565	148
2Nm (Hybrid to *F. polyctena*)	4.32E-05	-	5.38E-05
2Nm (*F. polyctena* to Hybrid)	0.5985248	-	4.2828787
2Nm (*F. aquilonia* to Hybrid)	0.5967169	3.9017825	-
2Nm (Hybrid to *F. aquilonia*)	3.5025E-08	0.0188780	-
2Nm (*F. aquilonia* to *F. polyctena*)	0.0035626	0.3522846	0.2855615
2Nm (*F. polyctena* to *F. aquilonia*)	4.61E-08	3.34E-08	0.0027451
Ancestral 2Nm (*F. aquilonia* to *F. polyctena*)	1.2082687	0.3545135	0.5127609
LogLikelihood	-888,805.413	-887,212.02	-887,116.789
ΔLikelihood	5,333.845	3,740.448	3,645.221

Supplementary Table 15 - Maximum likelihood parameter estimates for the Admixture models concerning the origin of the hybrid populations, tested with the dataset containing Långholmen R and the Finnish *Formica polyctena* and *Formica aquilonia* parental populations. All effective sizes (Ne) are given in number of haploids. Times are given in number of generations. Migration rates are scaled according to population effective sizes (2Nm). Maximum-likelihood estimates for parameters are taken from the run reaching the highest composite likelihood of the 100 runs performed. Likelihoods are given in logarithmic scale. Maximum observed likelihood for this dataset is -883,471.568. ΔLikelihood is calculated by subtracting the expected likelihood from the maximum observed likelihood.

Parameter	Admixture	Admixture with continuous migration
Ancestral N_e	425,785	414,391
***F. polyctena* ancestral N_e**	218,804	251,384
***F. aquilonia* ancestral N_e**	417,311	406,529
"Older" *F. polyctena* N_e	277,860	290,073
"Older" *F. aquilonia* Ne	215,480	202,620
Recent *F. polyctena* Ne	128	108
Recent Hybrid N_e	145	124
Recent *F. aquilonia* N_e	20,171	31375
Time of divergence	222,010	232,374
Time of admixture	48	41
Time of size change	201,597	171,429
Genetic input from *F. polyctena*	0.6508694	0.6498797
Genetic input from *F. aquilonia*	0.3491306	0.3501203
2Nm (Hybrid to *F. polyctena*)	-	3.18E-08
2Nm (*F. polyctena* to Hybrid)	-	1.80E-07
2Nm (*F. aquilonia* to Hybrid)	-	1.09E-03
2Nm (Hybrid to *F. aquilonia*)	-	1.12E-03
Ancestral 2Nm (*F. aquilonia* to *F. polyctena*)	0.770525	0.9022162
LogLikelihood	-886,877.549	-886,872.143
ΔLikelihood	3,405.981	3,400.575

Supplementary Table 16 - Maximum likelihood parameter estimates for the "Single Origin" model, tested with all datasets. All effective sizes (Ne) are given in number of haploids. Times are given in number of generations. Migration rates are scaled according to population effective sizes (2Nm). Maximum-likelihood estimates for parameters are taken from the run reaching the highest composite likelihood of the 100 runs performed. Likelihoods are given in logarithmic scale. ΔLikelihood is calculated by subtracting the expected likelihood from the maximum observed likelihood.

Parameter	Långholmen W + Långholmen R	Bunkkeri + Långholmen W	Pikkala + Långholmen W	Bunkkeri + Pikkala
Ancestral Ne	404,352	406,691	406,636	405,068
***F. polyctena* ancestral Ne**	279,281	208,897	230,995	220,531
***F. aquilonia* ancestral Ne**	322,204	428,417	463,201	478,486
"Older" *F. polyctena* Ne	259,630	183,460	217,564	167,326
"Older" *F. aquilonia* Ne	202,592	221,971	214,824	223,294
"Ancestral" hybrid Ne	94	865	1,261	1,579
Recent *F. polyctena* Ne	40	105	98	120
Recent Hyb1 Ne	21	163	134	251
Recent Hyb2 Ne	18	100	120	171
Recent *F. aquilonia* Ne	32,534	345	386	353
Time of divergence	271,601	293,771	288,905	296,577
Time of size change	256,974	290,220	286,032	291,931
Time of admixture	16	40	37	45
Time of divergence (Hyb1/Hyb2)	4	27	34	43
Genetic input from *F. polyctena* to Hyb. Ancestral	0.6846918	0.5820877	0.6249366	0.5803086
Genetic input from *F. aquilonia* to Hyb. Ancestral	0.3153082	0.4179123	0.3750634	0.4196914
Ancestral 2Nm (*F. aquilonia* to *F. polyctena*)	1.4875566	1.4897768	1.5956263	1.6074507
Max. Observed Likelihood	-693,262.061	-1,291,308.181	-1,001,623.462	-1,509,459.108
Expected Likelihood	-697,084.18	-1,294,945.22	-1,004,297.98	-1,511,635.29
ΔLikelihood	3,822.123	3,637.041	2,674.521	2,176.186

Supplementary Table 17 - Maximum likelihood parameter estimates for the "Independent Origins" model, tested with all datasets. All effective sizes (Ne) are given in number of haploids. Times are given in number of generations. Migration rates are scaled according to population effective sizes (2Nm). Maximum-likelihood estimates for parameters are taken from the run reaching the highest composite likelihood of the 100 runs performed. Likelihoods are given in logarithmic scale. ΔLikelihood is calculated by subtracting the expected likelihood from the maximum observed likelihood.

Parameter	Långholmen W + Långholmen R	Bunkkeri + Långholmen W	Pikkala + Långholmen W	Bunkkeri + Pikkala
Ancestral Ne	421,058	404,781	418,964	404,423
***F. polyctena* ancestral Ne**	285,725	232,964	216,549	217,188
***F. aquilonia* ancestral Ne**	342,327	369,004	325,280	355,582
"Older" *F. polyctena* Ne	196,118	195,966	193,630	192,074
"Older" *F. aquilonia* Ne	176,855	194,961	194,435	195,325
***F. polyctena* Ne**	224,301	237,725	227,749	185,924
***F. aquilonia* Ne**	212,286	58,709	157,543	111,357
Recent *F. polyctena* Ne	85	97	102	111
Recent Hyb1 Ne	157	221	188	218
Recent Hyb2 Ne	124	134	123	159
Recent *F. aquilonia* Ne	29,437	546	440	346
Time of divergence	250,956	264,120	277,720	231,839
Time of size change	221,407	256,045	264,695	218,637
Time of admixture (Hyb1)	38	38	48	40
Time of admixture (Hyb2)	34	37	36	41
Genetic input from *F. polyctena* to Hyb1	0.7662315	0.6195929	0.615766	0.6012129
Genetic input from *F. aquilonia* to Hyb1	0.2337685	0.3804071	0.384234	0.3987871
Genetic input from *F. polyctena* to Hyb2	0.7804881	0.676826	0.6352043	0.6149353
Genetic input from *F. aquilonia* to Hyb2	0.2195119	0.323174	0.3647957	0.3850647
Ancestral 2Nm (*F. aquilonia* to *F. polyctena*)	2.0293141	1.6793124	1.4669882	1.4914614
Max. Observed Likelihood	-693,262.061	-1,291,308.181	-1,001,623.462	-1,509,459.108
Expected Likelihood	-699,172.90	-1,295,106.47	-1,004,396.69	-1,511,762.49
ΔLikelihood	5,910.837	3,798.286	2,773.232	2,303.385