The first steps toward a global pandemic: Reconstructing the demographic history of parasite host switches in its native range

Maeva A. Techer[1]*, John M. K. Roberts[2], Reed A. Cartwright[3,4], and Alexander S. Mikheyev[1,5]*

1 Okinawa Institute of Science and Technology, 1919-1 Tancha Onna-son, 904-0495 Okinawa, Japan
2 Commonwealth Scientific & Industrial Research Organisation, Canberra, ACT, 2601, Australia
3 The Biodesign Institute, Arizona State University, Tempe, Arizona, USA
4 School of Life Sciences, Arizona State University, Tempe, Arizona, USA
5 Australian National University, Canberra, ACT, 2600, Australia

Corresponding Authors *

Maeva A. Techer maeva.techer@oist.jp, Alexander S. Mikheyev sasha@homologo.us

Maeva A. Techer 0000-0001-5417-5103
John M. K. Roberts 0000-0001-9739-5595
Reed A. Cartwright 0000-0002-0837-9380
Alexander S. Mikheyev 0000-0003-4369-1019

Keywords

Host switch, population genomics, demographic history, founder size, sympatry.

Abstract

Host switching allows parasites to expand their niches. However, successful switching may require suites of adaptations and may decrease performance on the old host. As a result, reductions in gene flow accompany many host switches, driving speciation. Because host switches tend to be rapid, it is difficult to study them in real time and their demographic parameters remain poorly understood. As a result, fundamental factors that control subsequent parasite evolution, such as the size of the switching population or the extent of immigration from the original host, remain largely unknown. To shed light on the host switching process, we explored how host switches occur in independent host shifts by two ectoparasitic honey bee mites (*Varroa destructor* and *V. jacobsoni*). Both switched to the western honey bee (*Apis mellifera*) after it was brought into contact with their ancestral host (*Apis cerana*), ~70 and ~12 years ago, respectively. *Varroa destructor* subsequently caused worldwide collapses of honey bee populations. Using whole-genome sequencing on 44 mites collected in their native ranges from both the ancestral and novel hosts, we were able to reconstruct the known temporal dynamics of the switch. We further found that hundreds of haploid genomes were involved in the initial host switch, and, despite being greatly reduced, some gene flow remains between mites adapted to different hosts remains. Our findings suggest that while reproductive isolation may facilitate fixation of traits beneficial for exploitation of the new host, ongoing genetic exchange may allow genetic amelioration of inbreeding effects.

Introduction

Arms races between parasites and their hosts drive evolutionary innovation. Novel parasites can decimate host populations or drive them extinct, unless counter-adaptations evolve. Similarly, parasite evolution accompanies the very act of host switching, since it requires adaptations to novel host physiology in order to persist and to spread. Because parasite adaptations tend to be host-specific, host switches are often associated with host-associated genetic differentiation and eventual speciation. However, only the endpoint of this process is typically observed, as host switches tend to occur rapidly, and the original host is often unknown. As a result, many unanswered questions remain about how parasites acquire new hosts. For instance, if host switches are accompanied by a bottleneck due to reduced gene flow from the ancestral host, how does the parasite have sufficient genetic diversity to adapt? Does gene flow cease completely, or does it continue at a low level, potentially providing additional genetic material for adaptations?

One of the major limiting factors for host switching is the geographic separation between parasites and potential hosts. Globalization has eased these barriers, sometimes giving rise to pandemics (1). As a result, host switches are easier to observe and to study in something approaching real time. One of the most dramatic and economically important switches involved the two ectoparasitic mite species, *Varroa destructor* and *Varroa jacobsoni*, which acquired the western honey bee (*Apis mellifera*) as a new host, ~70 and ~12 years ago, respectively (2, 3). *V. destructor,* in particular, spread worldwide, causing extensive honey bee population collapses, whereas *V. jacobsoni* has so far remained in Oceania (4, 5). Both mites were originally found on the sister species, *Apis cerana*, and came into contact with *A. mellifera,* which was brought in for purposes of beekeeping (6, 7). These two host switches occurred in parallel and relatively recently, allowing the reconstruction of how the host switches took place, using genomic tools.

Both switches have been investigated using microsatellite markers and mitochondrial markers, which revealed that in both species, populations parasitizing *A. mellifera* were strongly differentiated and genetically depauperate (3, 8–12). While *V. destructor*, in particular, was described as "quasi-clonal" (8), it successfully spread worldwide, and has shown a remarkable ability to parasitize genetically diverse *A. mellifera* populations, as well as to evolve resistance to human counter-measures, such as pesticides (13). How does a bottlenecked species achieve such a level of success? Increasing evidence from population genomic analysis of fungal pathogens suggests that the success of many pathogens appears to rely on maintaining some level of adaptive diversity despite the presence of bottlenecks during host switches (14, 15), but similar work is lacking in animal systems.

In order to answer this question and to gain broader insight into how host switches happen, we sequenced genomes from sympatric populations of the two mites across Asia and Oceania, collected on both novel and introduced hosts. This allowed us far greater power to examine how the host switch took place with much greater precision than was possible previously. We found strikingly parallel dynamics at play in both host switches, which were characterized by a surprisingly large effective population size at the time of the switch and ongoing gene flow with cryptic population genetic processes that may have helped *Varroa* succeed.

Results

We examined and compared demographic histories by sequencing whole genomes from 19 *V. destructor* and 24 *V. jacobsoni* diploid female mites sampled on their original host, *A. cerana,* and on the novel one, *A. mellifera* (Table 1). Sampling sites spanned the mites' native ranges corresponding to *A. cerana*'s native and extended range (Figure 1). In those areas, both parasitized hosts occur sympatrically from Northeast China (Heilongjiang province) to Papua New Guinea (details in Table S1). Additionally, six pairs of diploid mother-foundresses *V. destructor* and their haploid male offspring were sequenced to estimate non-inherited variability between generations. Paired-end Illumina reads were mapped against the *V. destructor* reference genome (16) with a mean read coverage from ~10x to ~50x (except for one mite from Thailand) (Table S2).

Table 1: Summary of sampling details of each *Varroa* specimen from either its original or new host.

Species	Host	Country	Date	N	mtDNA *COX1* (Anderson and Trueman 2000)
V. destructor	*A. cerana*	South Korea	1996	1	VD Korea K1
		China	2001-2002	6	VD China C1 VD China C3 VD China C4 VD Viet Nam V1
		Thailand	2003	1	VD Viet Nam V1
		Viet Nam	1996	1	VD Viet Nam V1
V. destructor	*A. mellifera*	South Korea	1996	2	VD Korea K1
		China	2001	5	VD Korea K1
		Viet Nam	1996, 1998	2	VD Korea K1
		Okinawa [a]	2018	12	VD Korea K1
V. destructor	Mis assigned host	Viet Nam	1996	1	VD Korea K1
V. jacobsoni	*A. cerana*	Viet Nam	2003-2004	4	VJ Laos L1
		Malaysia	1995	1	VJ Malaysia 2
		Indonesia	1998, 2002	6	VJ Java 1 VJ Lombok 2
		Papua New Guinea	1997, 2008, 2015	4	VJ PNG 1
V. jacobsoni	*A. mellifera*	Papua New Guinea	2008, 2014	8	VJ PNG 1
V. jacobsoni	Mis assigned host	Papua New Guinea	2014	1	VJ PNG 1

[a] *V. destructor* female mother-foundresses and sons used for estimating the mutation rate.

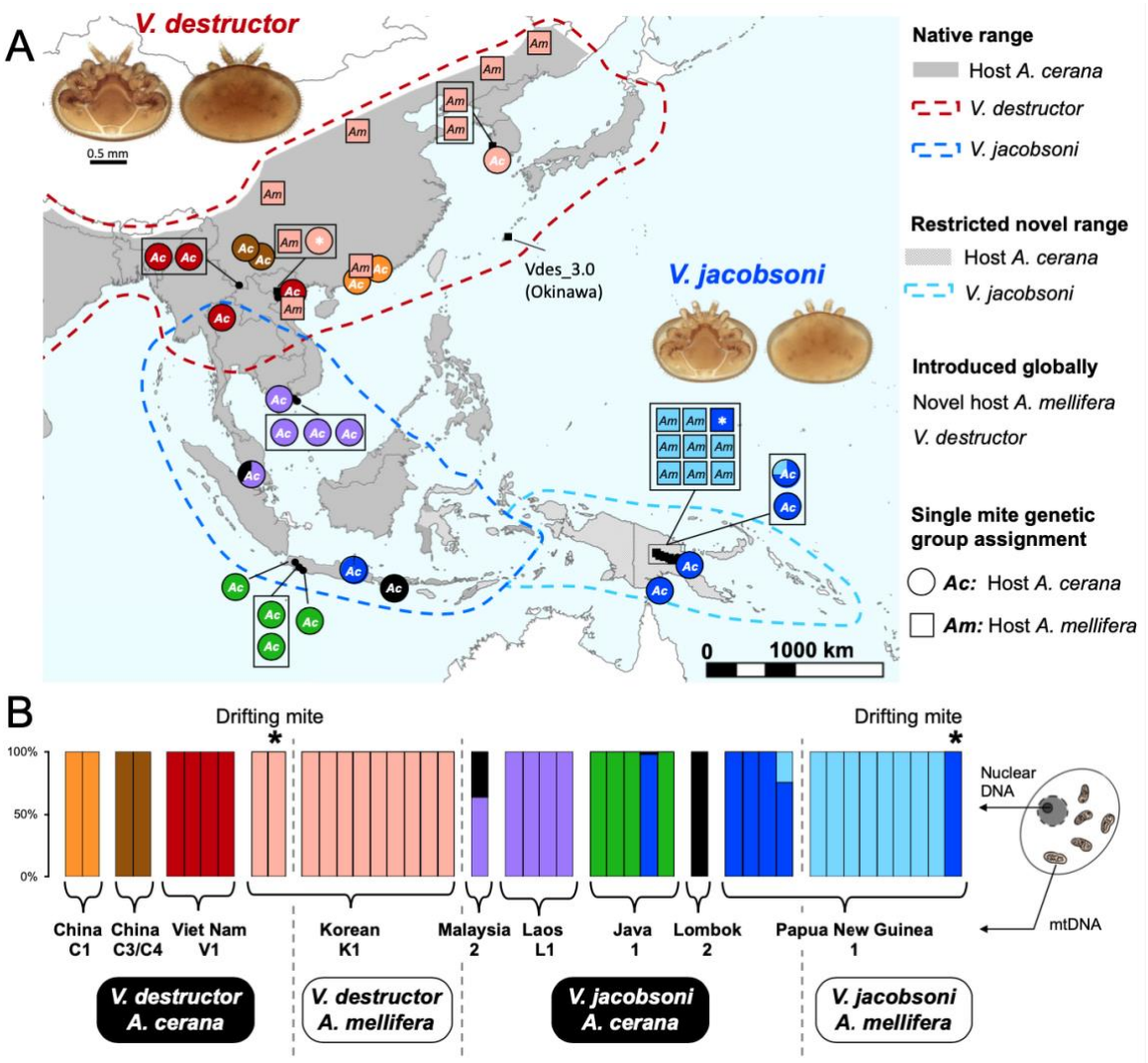

Figure 1. Populations of sister species, *V. destructor* and *V. jacobsoni,* are genetically structured within their native range and between honey bee hosts. A) Following *V. destructor* coevolution with its original host *A. cerana* (Ac), mite populations diverged across their native range (red-dotted line). In contrast, sympatric mite populations established on *A. mellifera* (Am) following host switching (~1950s) suffered from a founder effect and are genetically homogenous throughout Asia. Parallel evolutionary patterns are observed for the parapatric *V. jacobsoni* in South-Asia and Oceania. B) Mite ancestry and population structure, considering K = 9 groups matched with the distribution of mitochondrial defined haplogroups/lineages. In both species, presence of drifting mites (i.e., migrating from a honey bee colony to another using a flying host, and denoted as *) between sympatric hosts was directly evidenced by host read identity and genetic assignment analysis.

Genetic founder effect in Varroa mites' lineages

We confirmed species identity by aligning *Varroa* mitogenomes together with available reference sequences of the mitochondrial (mtDNA) *COX1* gene (458-bp) widely used for *Varroa* mites (17). The nucleotide variability detected in our consensus sequences with high coverage (cut-off at 100x) matched exactly the variability observed using the Sanger sequencing method. Species identification and parapatric distribution were confirmed with mtDNA *COX1* barcoding. Distinction of haplogroup and species relied on 39 segregating sites within the COX1 region, with a species divergence ranging from 5.9% - 7.0%. All *Varroa* mites (including families)

clustered within diverse known haplogroups, i.e., mitochondrial lineages (4), with the exception of a newly described haplogroup (VJ Lombok 2) (Figure S1 and Table S3). The haplogroup distribution observed in this study was consistent with the geographical origin reported for reference sequences used for comparison (4).

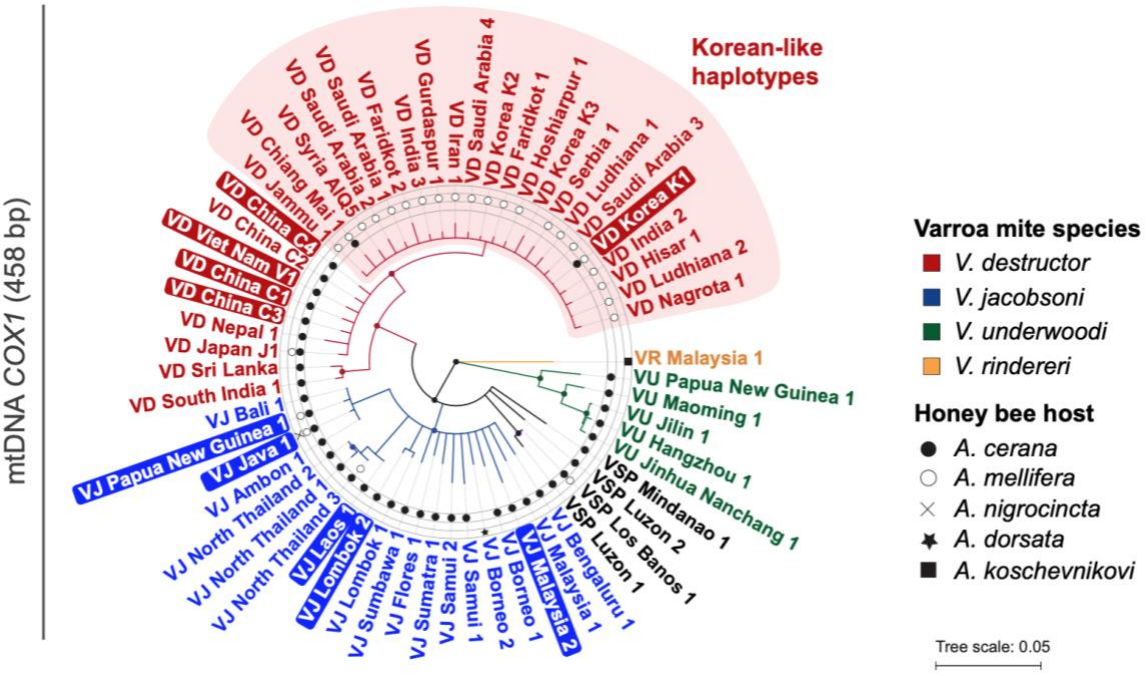

Figure S1: Sequenced *V. destructor* and *V. jacobsoni* mites originating from diverse maternal genetic backgrounds (haplogroups) as shown by mtDNA *COX1* barcoding. For *V. destructor* (red): mites collected throughout Asia on their original host belong to five distinct haplogroups: Korea K1, China C1, C3, C4 (ex China2), and Viet Nam V1 (see Table S3). By contrast, all the host-switched mites across that range belong to the Korean K1 haplotype [AF106899], which is also found worldwide (2). For *V. jacobsoni* (blue): mites collected throughout Southeast Asia and Oceania on *A. cerana* presented partial *COX1* sequences clustering within five haplogroups: Papua New Guinea 1, Java 1, Laos 1, Malaysia 2, and a new one named Lombok 2. In contrast, all *A. mellifera* were singly infected by one *V. jacobsoni* haplogroup, described as Papua New Guinea 1. A neighbor-joining tree was built using the Tamura-Nei genetic distance model with 1,000 bootstraps (< 80% are shown as filled circles in the base of each branch). *V. rindereri* was considered as an outgroup and the host range indicated (4).

The diversity of native mite haplogroups (based on the COX1 region) from *A. cerana* was severely reduced on the novel host, *A. mellifera*. Despite a similar number of mites sequenced per host throughout Asia and Oceania, we confirmed that exclusively *V. destructor* Korean K1-origin mites and *V. jacobsoni* Papua New Guinea 1-origin were found on *A. mellifera* (Figure 1). For *V. destructor*, the Korean K1 haplogroup is the same one suspected to have switched hosts initially in the Korean peninsula near Vladivostok in the late 1950s and now cosmopolitan (6, 10). Using mtDNA, we found no evidence of the less frequent Japanese J1 haplogroup, even in its native range (Okinawa), although it switched hosts in Japan, possibly in the 1950s (2, 10). For *V. jacobsoni*, the Papua New Guinea 1 haplogroup is also the same as that reported during the host switch event which occurred in 2008 (3). As a consequence, patterns of host specificity emerged especially for *V. destructor,* as mtDNA lineages on *A. cerana* were geographically structured, contrary to mites on *A. mellifera* which occur sympatrically (Figure 1A). Trends of mtDNA lineage specialization on *A. mellifera* were clearly noticeable in South China and North Viet Nam.

We scaled our mitogenome consensus sequences to compare them to the highest resolution available regarding *Varroa* intra-haplogroup genetic diversity (10). We found that more sequence variability exists within *V. destructor* mites on *A. mellifera* than previously reported based on the longest standard mtDNA fragment available on NCBI (2,696 bp from *COX1-ATP6-COX3-CYTB* concatenation) (Figure 2A). Two new Korean K1-derived haplotypes (*V. destructor*) were observed in *A. mellifera* mites in China, and one individual presented two haplotypes differing by one SNP (confirmed with sequence electropherogram as described for *Varroa* (18)). The star-like shape of the Korean K1 haplotype network suggests that either *V. destructor* on *A. mellifera* evolved novel mutations following host switch and range expansion or that additional jumps occurred in Asia from unsampled source populations. In contrast, for *V. jacobsoni*, we detected 11 mitochondrial haplotypes, but mites on *A. mellifera* exhibited only the Papua New Guinea 1-1 haplotype. This was true of mites collected at the time of the reported host switch event (2008) (3) and six years later (2014) (Figure 2B and Table S1, S2).

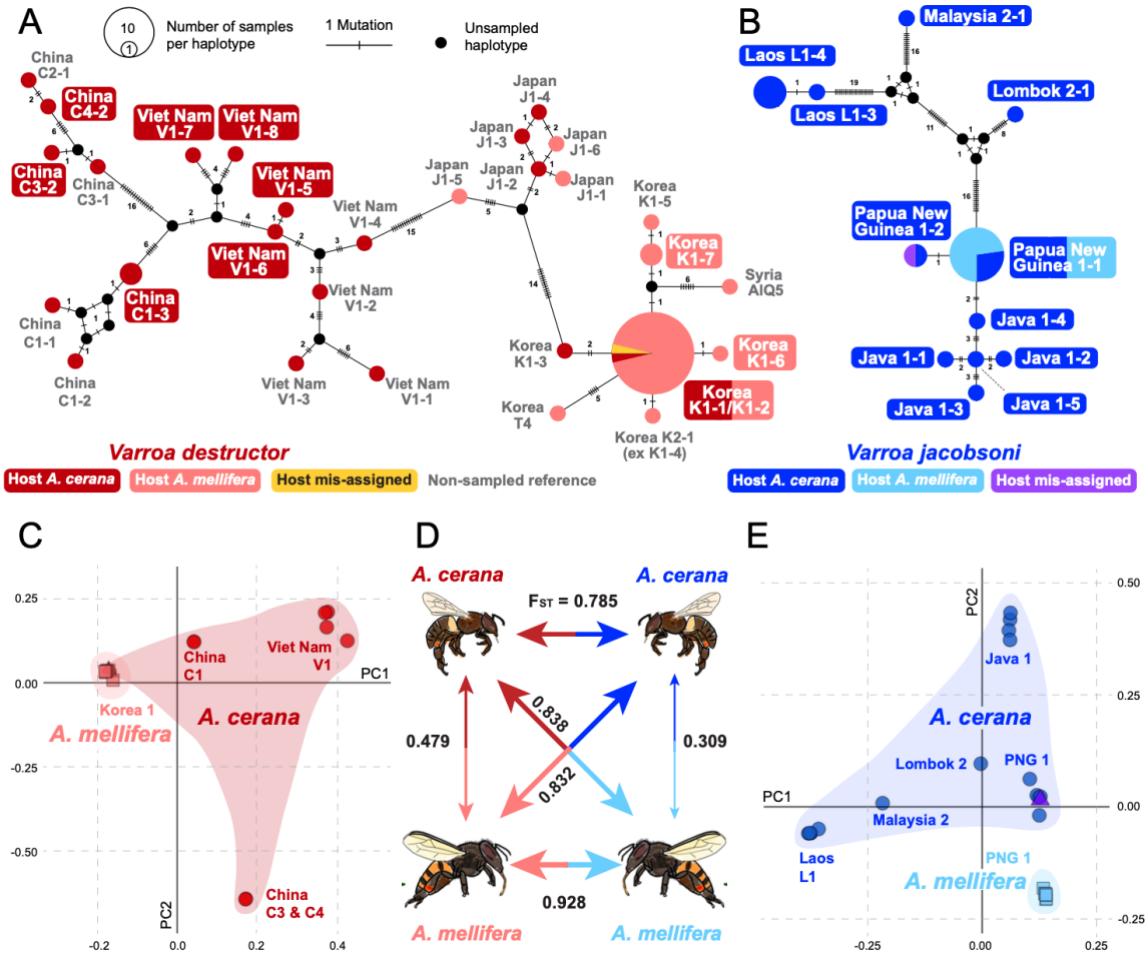

Figure 2: Loss of genetic diversity and rapid genomic differentiation occurred in spite of large founding size and migration between sympatric honey bee hosts. On their original host (A) *V. destructor* diverged into at least 20 maternal lineages, including seven newly described by this study and (B) *V. jacobsoni* diverged into 11 mitochondrial lineages, as seen on the *COX1-ATP6-COX3-CYTB* mtDNA median-joining haplotype network. Mites sequenced in this study encompassed various divergent lineages in both species (names highlighted in red or blue). Only a fraction of mite genetic diversity in their native range has become established on *A. mellifera,* despite similar opportunities. The star-shaped pattern observed for Korean K1 haplotypes could indicate additional jumps from unsampled source populations and/or population expansion on *A. mellifera*. The Korean K1-origin mites successfully established on *A. mellifera* genetically differentiated from all other *A. cerana*

populations, as shown on a PCA based on 198,558 pruned SNPs (C) with high genome-wide weighted F_{ST} indices (D). Similarly, *V. jacobsoni* colonizers from *A. mellifera* also rapidly differentiated from the sympatric *A. cerana* source population in Papua New Guinea (PNG1) since its host switch in 2008 (E).

Genetic diversity and differentiation within and among host populations

Our variant calling from 44 *Varroa* nuclear genomes (368Mb, seven chromosomes) revealed that 4,056,305 genomic sites were variable. Among these, we detected 1,471,976 biallelic SNPs for which the second allele frequency was ≥5% throughout the whole dataset (minor allele frequency MAF > 0.05). *V. destructor* and *V. jacobsoni* were differentiated in their allele frequency distribution as well as at the intra-species level by considering mtDNA lineages. Host switched populations on *A. mellifera* exhibited much lower genome-wide variability than those from *A. cerana* populations (see nucleotide diversity π computed on 2,604,602 sites, including monomorphic ones, Figure S2). Nucleotide diversity on all seven chromosomes was inversely correlated with the time since host switch. On average a 6-fold difference was observed for *V. destructor* (that switched hosts in ~1950s) versus a 3-fold reduction in *V. jacobsoni* that jumped in 2008.

Nonetheless, *V. destructor* on *A. mellifera* differentiated over 332,845 biallelic SNPs while standard *Varroa* mtDNA markers did not show such differences. Yet, 75.6% of the variability (n = 81,361 biallelic SNPs) among *V. destructor* mites on *A. mellifera* was contributed by rare alleles found in <10% of this host population in Asia. Similarly, we observed alike patterns for *V. jacobsoni* mite genomes on *A. mellifera* that segregated over 349,894 biallelic SNPs, in spite of the loss of genetic diversity compared to *A. cerana* mites on geographically adjacent sites in Papua New Guinea.

Following linkage disequilibrium depletion of the SNPs dataset, we observed similar evolutionary trends using 198,558 SNPs distributed over the seven mapped chromosomes. Colonization and evolution of *Varroa* mites on the novel host was not associated with significant differences in inbreeding levels between original and novel host populations (Table S4). We estimated high levels of genome-wide inbreeding coefficients for *V. destructor* on *A. cerana* (average F_{IS} = 0.858 ± 0.075), comparable to those of *A. mellifera* mites (average F_{IS} = 0.892 ± 0.053). Extremely low levels of observed genomic heterozygosity in novel host *V. destructor* populations (mean H_{obs} = 0.03 ± 0.01) were comparable to earlier reports in other Asian populations present in Thailand (mean H_{obs} = 0.04 ± 0.01 and 0.05 ± 0.01 using microsatellites (12)) or China and Japan (8). In parallel, *V. jacobsoni* also exhibited high levels of inbreeding in both populations, but with variability among maternal lineages in *A. cerana* (average F_{IS} = 0.683 ± 0.203) and within *A. mellifera* Papua New Guinea mites (average F_{IS} = 0.754 ± 0.131). Comparatively, heterozygote proportions were within the range reported in the same geographical regions in a microsatellite study (3). Such low levels in both species, regardless of host population, are more likely attributed to incestuous *Varroa* mating rather than to a host switch founder effect alone.

Different methods (PCA, Bayesian clustering method, weighted F_{ST} and D_{xy}) converged and indicated many levels of genetic structure with divergence across genomes: 1) between *Varroa* species, 2) within and between *A. cerana* and *A. mellifera,* 3) geographically associated with mtDNA lineages. Extremely high F_{ST} levels (Table S5) and genetic differentiation occur between hosts within *Varroa* species (Figure 2CDE). Within species, genetic divergence increased among divergent mtDNA lineages, but most interestingly within lineages between hosts, as shown with *V. jacobsoni* mites originating from Papua New Guinea 1-origin mites (Figure 2CE).

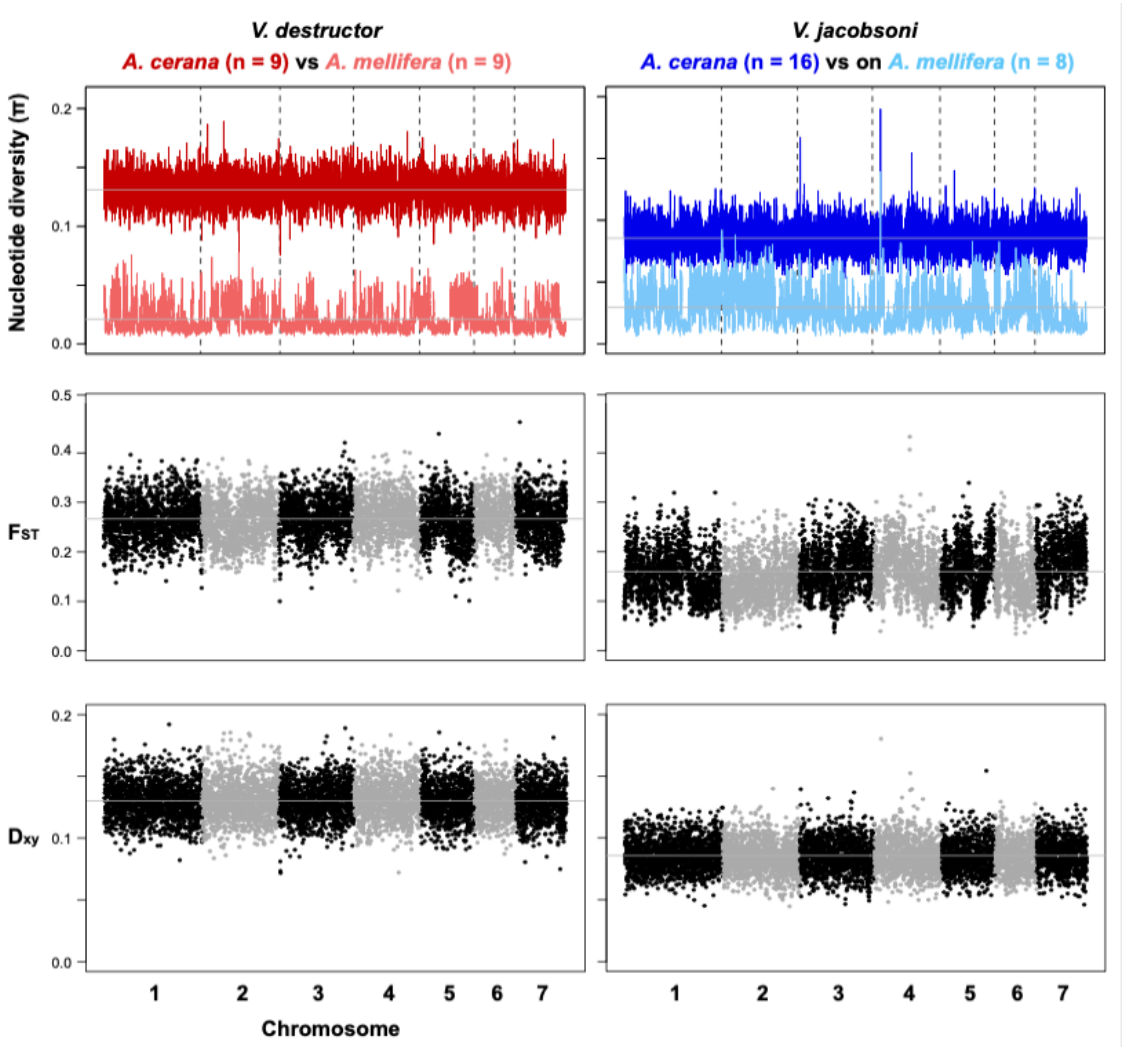

Figure S2: Parallel reduction of genetic diversity and genomic divergence between hosts in *V. destructor* and *V. jacobsoni*. First row: Nucleotide diversity in 50kb window along the seven chromosomes with all mites parasitizing the same host species pooled as one population for each *Varroa*. Genetic divergence estimated by pairwise F_{ST} (second row) and absolute divergence using D_{xy} estimate (third row) in 50kb window, show a strong host differentiation all along the genome with few outliers.

Evidence for migrating/drifting mites from genomic ancestry and host reads

We applied an incremental step method to select an NGS K model that would be coherent with differentiation patterns observed while detecting potential admixture among mite lineages. The model with nine genetically distinct clusters was the most biologically sound, and reflected the population structure detected of *A. cerana* populations in our other analysis (Figures 1B and S3). Using *A. cerana* mites as reference populations, we found no evidence of ancestry other than Korean K1 mites for *V. destructor* and Papua New Guinea 1 for *V. jacobsoni* (Table S6). The absence of introgression of lineages could not simply be attributed to geographic distance, especially for mites in Guangdong, China and North Vietnam (Figure 1B), where various mtDNAs co-exist on different hosts. One *V. jacobsoni* specimen in Indonesia (Gunung Arca) appeared as an outlier among A. *mellifera* mites and clustered with Papua New Guinea mites on *A. cerana*. The high divergence of this individual from other mites on *A. mellifera* suggested a recent migration event.

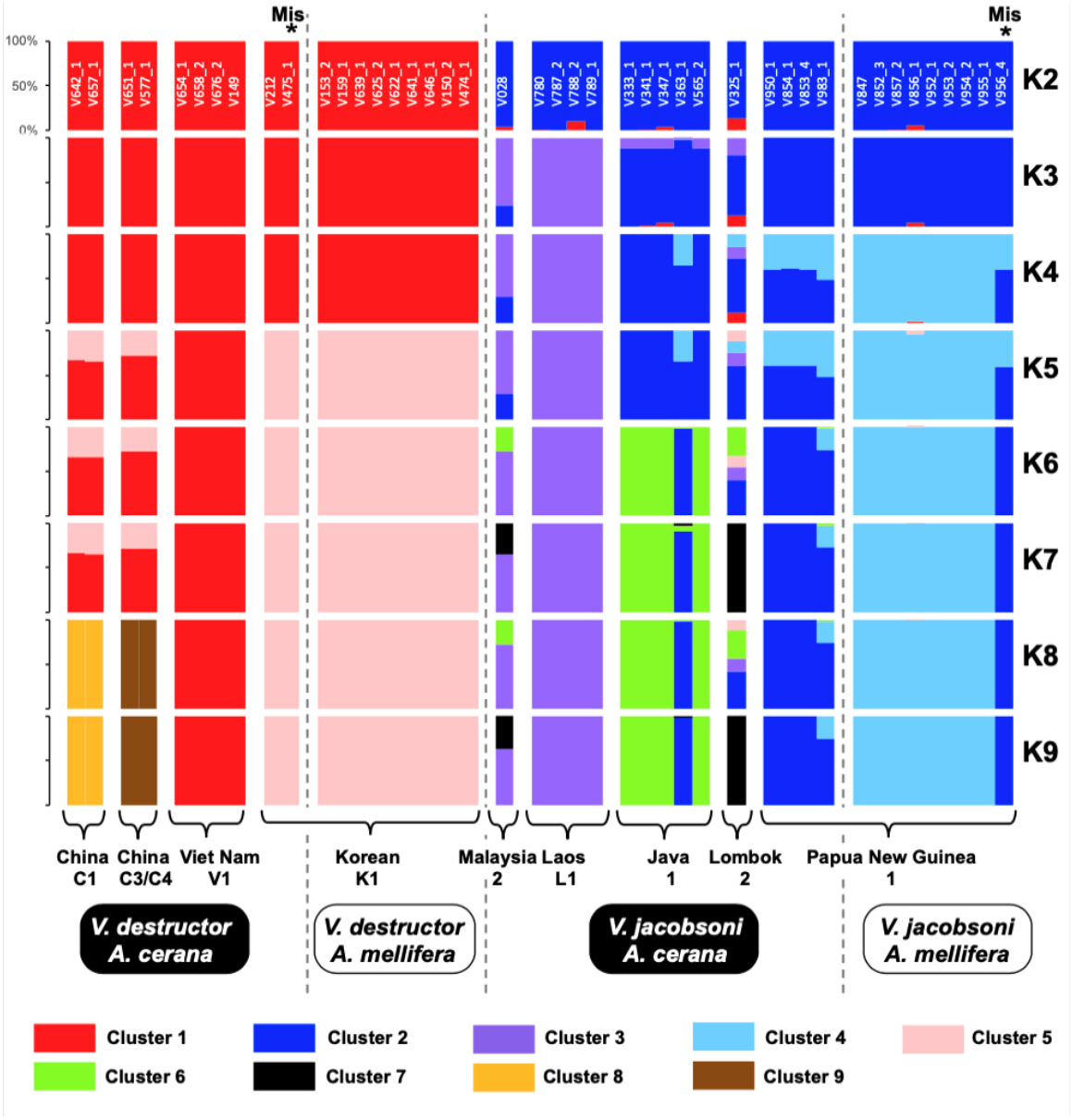

Figure S3: Population structure of *Varroa* mites in their native areas when choosing the best genotype likelihood run for each K model. The choice of K = 9 as a model was the most biological probable to follow species, host, mitochondrial lineages and geographical structure trends observed. Clustering was computed with no population priors.

Recent migration and spillover events were confirmed by analyzing mite's metagenome (gut contents). Given the parasitic nature of *Varroa* mites and the library preparation by crushing whole mites, it was assumed that mitochondrial reads from its hosts could be found. Indeed, after read quality score filtering, up to 0.18% of all reads (n = 115,329 reads) matched honey bee host mtDNA (Table S2). The proof of concept for this method was first tested and demonstrated with mite families that were actively feeding at the time of collection. Interestingly, we confirmed that the genetic outlier *V. jacobsoni* specimen (V956_4) in *A. mellifera* (Figure 1B) presented more *A. cerana* reads (n = 1,444) than of the collected host *A.*

mellifera (n = 69). While NGSadmix analysis did not reveal such an outlier for *V. destructor*, the diet molecular analysis revealed that one mite (V475_1) showed an imbalance of host reads in favor of *A. mellifera* (n = 555 reads) instead of the host *A. cerana* identified in the field (n = 6). Given the obvious morphological differences existing between *A. mellifera* and *A. cerana*, and the expertise of field collectors, such convergent observations more likely indicate mites that migrate between hosts shortly before collection.

Low mutation rate estimated from pedigrees

We examined whether the rapid parallel differentiation observed between *Varroa* host populations was driven by high frequency of new mutations by estimating the *de novo* mutation rate. As the *Varroa* reproductive sequence cycle is well known (one haploid son, followed by up to four diploid daughters), we sequenced six diploid mothers and their respective sons at a coverage between 18x - 33x. Before filtering, the number of detected mutations was 3,082 but there was extreme variation between samples in the number of mutations called. Notably mother-son pairs with less coverage contained several fold more mutation calls than mother-son pairs with higher coverage. Samples with higher mutation calls also showed stronger evidence of clustering of mutation calls, which typically indicates false-positive mutation calls. After filtering, two mutation calls remained, but showed low coverage and were discarded as unreliable.

Out of 6,000 simulated mutations, 3,464 could be detected by our pipeline, indicating that our recall rate was approximately 58%. The total haploid length of the 7 chromosomes considered in the mutation scan was 362,833,908 base pairs. Given that we found no mutations among 6 mother-son pairs, we estimate that the mutation rate in *Varroa* is less than 8×10^{-10} per bp per generation ($\frac{1}{7 \times 362833908 \times 0.5773}$) or less than 0.28 mutations per son per generation.

Demographic history inferences at and after each host switch.

We tested whether the size of randomly LD-pruned SNPs affects the best evolutionary scenario choice by running independently demographic inferences on four different subsets, and this for each of the *Varroa* species (*V. destructor* datasets VD1 = 377; VD2 = 3,760; VD3 = 12,665; VD4 = 38,108 SNPs; *V. jacobsoni* datasets VJ1 = 472; VJ2 = 4,713; VJ3 = 15,363; VJ4 = 46,416 SNPs). Note that two host mis-assigned host samples were not included, and population pairs composition assumed the biological possibility that *Varroa* mites could access any of the hosts in sympatry or via beekeeping exchanges. Given the low population connectivity observed in Indonesia populations for *V. jacobsoni*, we only included sympatric mites found either on *A. cerana* or *A. mellifera* from Papua New Guinea. In each of the fastsimcoal2 run per *Varroa* species, the same scenarios were consistently ranked as the most likely, based both on the Akaike Index Criterion (AIC) and maximum likelihood distribution from pseudo-observed datasets (Table S7 and Figure S4).

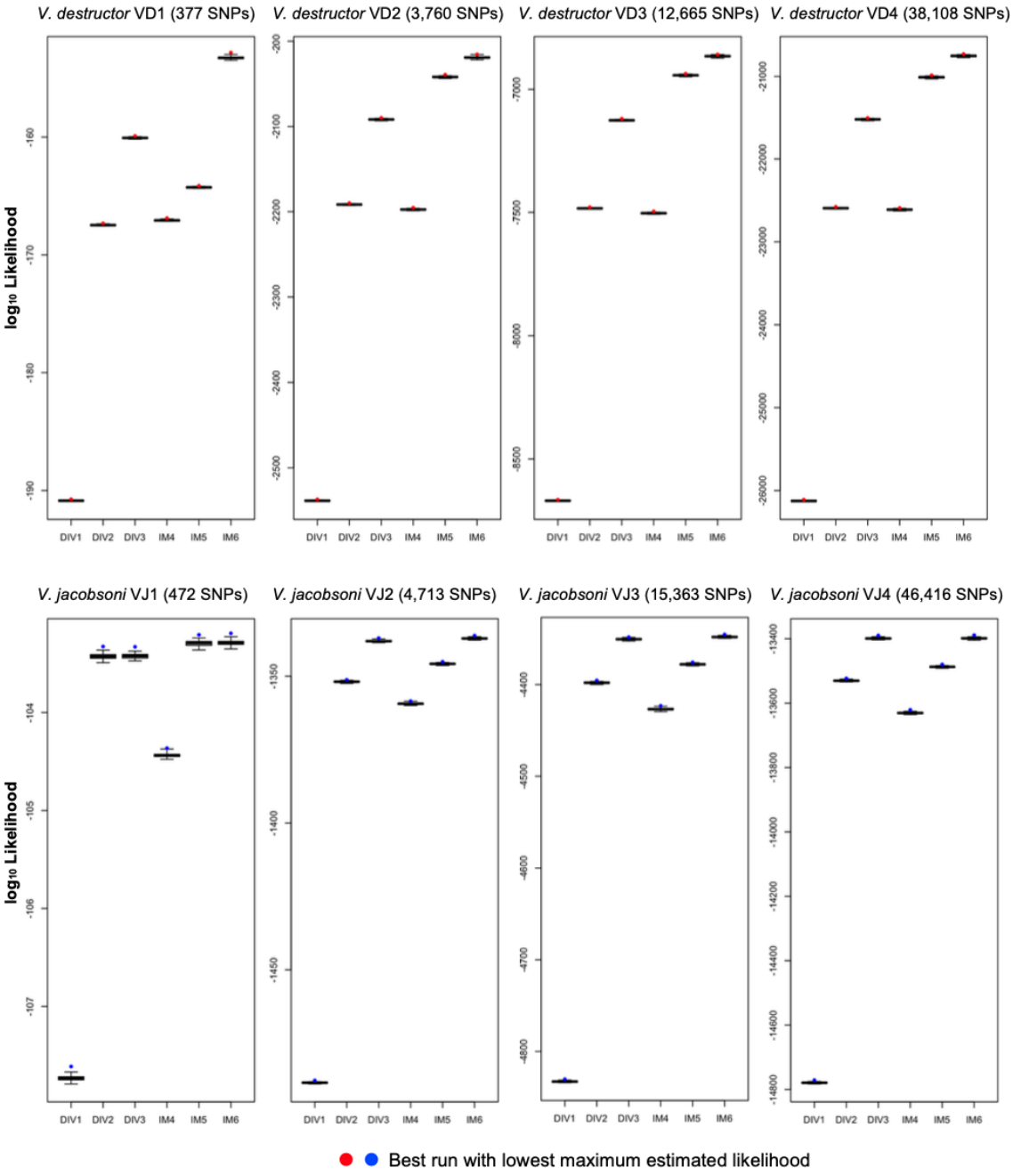

Figure S4: Maximum likelihood distribution for best scenario choice using the four datasets for *Varroa* mite species. The most likely scenario has the lowest value with scenario 6 for *V. destructor* and scenario 3 and 6 for *V. jacobsoni.*

As ranking of demographic models appeared consistent among site frequency spectrums (SFS) generated from different SNPs datasets, we selected the most likely scenario using the largest datasets (VD4 and VJ4). For *V. destructor*, pseudo-observed SFS simulated under a scenario with divergence between hosts accompanied by bi-directional asymmetric gene flow and population expansion on *A. mellifera*, was the most likely to fit the actual SFS (AIC = 41473.3 and weighted AIC ω = 1.00, Figure 3). The second best model was a scenario for which gene flow followed a

much stricter bottleneck with only one founder diploid mite. Comparison of observed and expected SFS under the best scenario, deviated least from the demographic model (Figure S5). For *V. jacobsoni*, weighted and unweighted Akaike criteria suggested scenario 3 (AIC = 26789.7 and ω = 0.750) and 6 (AIC = 26791.9 and ω = 0.250) as most likely to fit the observed SFS (Table S7 and Figure 4). The overlapping distribution of log $_{10}$ likelihood of pseudo-observed SFS generated by these best scenarios does not favor one model over the other (Figure 3). Both demographic models share divergence among hosts populations following a recent split from *A. cerana* and a population expansion either with migration (scenario 6) or without (scenario 3).

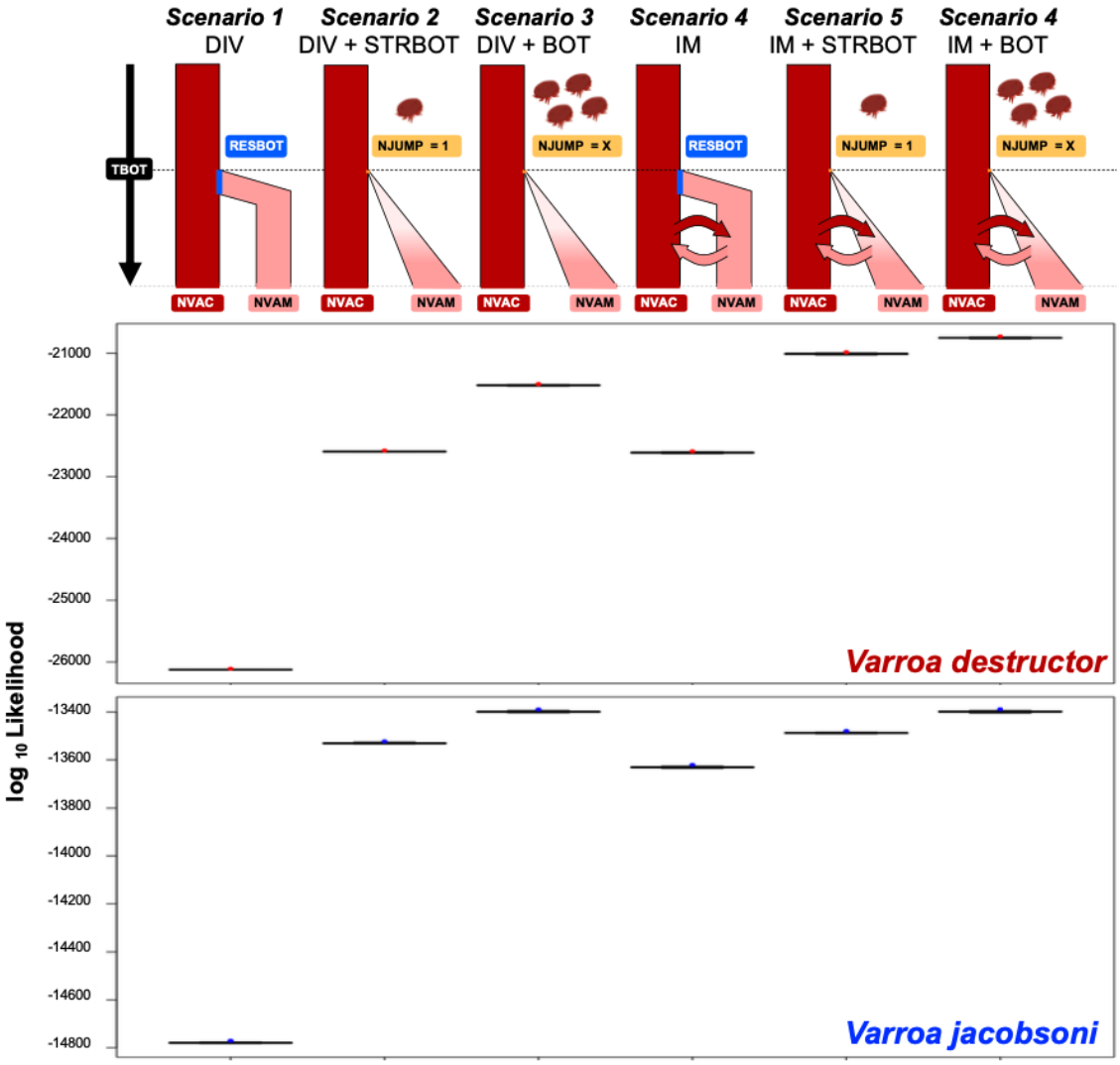

Figure 3. Best scenarios considered either a medium (BOT) or strong founder effect (STRBOT) associated with host switch, followed by migration between hosts (IM) for *V. destructor*. Model selection and comparison suggest similar scenarios for *V. jacobsoni* with or without migration events. For each species, log $_{10}$ likelihood distributions in boxplots based on 100 pseudo-observed SFS datasets generated from the best point estimates in each scenario.

Once the best model was selected, we estimated the host switch and subsequent population history parameters that could explain the observed SFS in *A. mellifera* mite populations. Our estimates suggest large founding sizes, on the order of hundreds of haploid genomes (Table 2) for both sister species, and regardless of whether ongoing migration was considered for *V. jacobsoni*. Assuming 10 mite generations per year, host switching has been traced to a time close to the reported host switch. Since their establishment on A. *mellifera*, both mite populations expanded with a growth rate for *V. jacobsoni* inferred to be about 18x (scenario 3) to 35x (scenario 6) of estimates for *V. destructor*. One interesting aspect of our demographic models was the continuously asymmetrical migration rates observed, suggesting frequent spillover from *A. cerana* to *A. mellifera* (Figure 4). All estimates of parameter estimates mean values were within our confidence intervals computed on 1,000 bootstraps simulations.

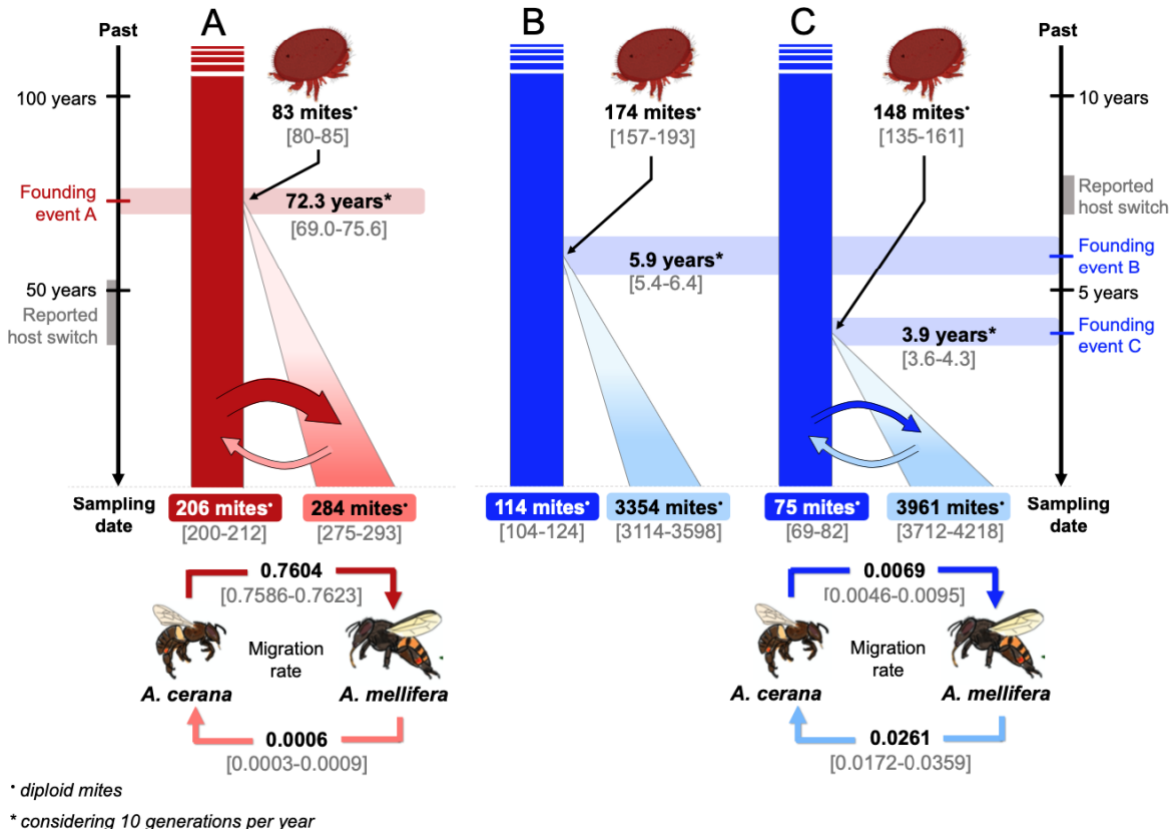

Figure 4: Multiple jumps on *A. mellifera* and isolation with migration likely allowed *V. destructor* and *V. jacobsoni* to successfully switch hosts and persist. Graphical illustration of the most likely scenario with best point parameter estimates in bold and associated 95% confidence interval in [light gray] with (A) isolation with migration for *V. destructor* in *A. mellifera*, (B) isolation with no migration and (C) with migration for *V. jacobsoni*. Estimated population sizes are given in haploid size and were transformed into diploid female mite units. Similarly, time estimates were given in generation time, but were represented here considering an average of 10 generations a year (53). Adjacent to the time arrow, grey areas indicate possible reported times of host switch events from the sampling date. Migration rates were kept as default values in haploid individuals per generation.

Table 2: Parameter estimates for the most likely demographic scenarios with isolation with migration and population expansion for each species jump.

	V. destructor			*V. jacobsoni*		
	Best point estimate	Mean value	CI 95%	Best point estimate	Mean value	CI 95%
NAmel (haploid individuals)	429	**568.1**	[550.3; 585.8]	8811	**7922.4**	[7424.0; 8436.0]
NAcer (haploid individuals)	229	**412.5**	[400.5; 424.7]	157	**150.5**	[137.6; 164.2]
TBOT (generation)	479	**723.0**	[689.8; 755.8]	41	**39.0**	[35.7; 42.5]
NJUMP (haploid individuals)	383	**165.0**	[160.4-169.8]	235	**295.1**	[269.6-322.0]
GAM (haploid individuals per generation)	5.7 x 10^{-03}	**3.4 x 10^{-03}**	[3.6 x 10^{-03}; 3.2 x 10^{-03}]	2.0 x 10^{-01}	**2.5 x 10^{-01}**	[2.8 x 10^{-01}; 2.3 x 10^{-01}]
NAmelMAcer (haploid individuals per generation)	4.2 x 10^{-05}	**5.9 x 10^{-04}**	[3.0 x 10^{-04}; 9.0 x 10^{-04}]	1.3 x 10^{-02}	**2.6 x 10^{-02}**	[1.7 x 10^{-02}; 3.6 x 10^{-02}]
NAcerMAmel (haploid individuals per generation)	7.6 x 10^{-01}	**7.6 x 10^{-01}**	[7.6 x 10^{-01}; 7.6 x 10^{-01}]	1.5 x 10^{-02}	**6.9 x 10^{-03}**	[4.6 x 10^{-03}; 9.5 x 10^{-03}]

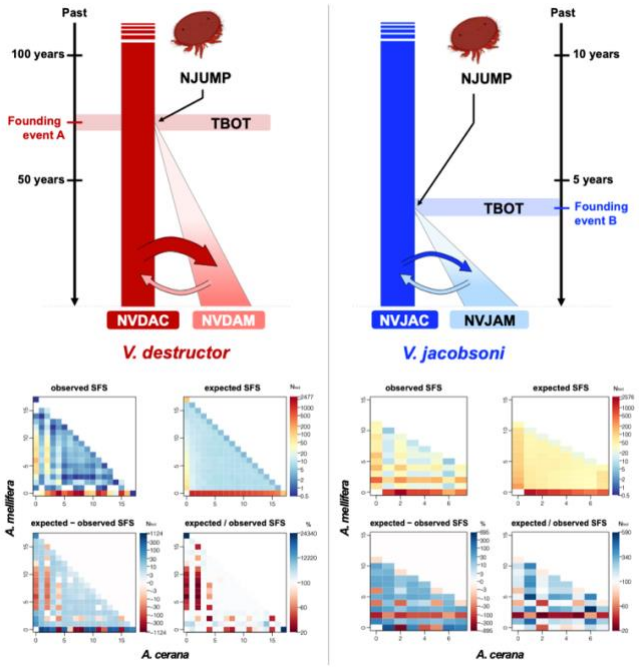

Figure S5: Observed vs expected joint 2D-SFS from the best scenario 6 including isolation with migration for *V. destructor* and *V. jacobsoni*

Discussion

Virtually every living species has at least one species of parasite, making parasitism perhaps the most successful mode of life. Most studies examine parasitism from either a macroevolutionary perspective, for instance, through host-parasite co-phylogenies, or at the population level, by examining patterns of differentiation between parasites and hosts (19, 20). Theoretical modelling links the two scales, suggesting that specialization coupled with trade-offs in performance on alternative hosts should lead to speciation (21). Yet, empirical observations of this process remain scarce, particularly of host switch demography, which includes key parameters such as population sizes during switches that are necessary to ensure sufficient evolutionary potential for parasites (22). For this reason, pandemic spread of *Varroa* mites among honey bees has long been puzzling, given their "quasi-clonal" structure in the invasive portion of their range (8). Using high-resolution genomic data from sympatric populations, we show that in both *V. destructor* and *V. jacobsoni,* (a) genetic bottlenecks were far less severe than previously estimated and (b) while gene flow was greatly reduced post-host switch, consistent with models of incipient speciation that may accompany acquisition of novel hosts, gene flow has not completely stopped. Our data highlight the importance of genetic diversity during initial stages of parasitic host switches.

While loss of genetic diversity is common in parasites, it often occurs in the parasite's introduced geographic range (23, 24). Consequently, parasites may possess sufficient genetic diversity to parasitize their hosts. For example, the fungal parasite responsible for ash dieback in Europe arrived from Asia, and while bottlenecked to only two haplotypes, it nonetheless maintains adaptive diversity in key host interaction genes (15). To the best of our knowledge, our study is unique in quantifying demographics within the native range of the parasite, where it has opportunities for additional gene flow from genetically diverse sympatric populations. At least for *Varroa,* the initial host switch requires an unexpected amount of genetic diversity at the point of the switch, but rapidly leads to reproductive isolation between sympatric populations on novel and original hosts. The amount of differentiation appears to increase over time, being greatest in *V. destructor*, which switched at least 50 years earlier than *V. jacobsoni*. *Varroa* mites coevolve quickly with their new hosts (25) and such sympatric isolation could ultimately lead to speciation.

Post host-switch speciation can occur in the presence of gene flow, if specialization is adaptively favored and selection acts on many genomic regions (26, 27). By promoting genomic heterogeneity and introducing beneficial mutations (28), gene flow can ameliorate lost diversity due to selection or inbreeding (29, 30), both of which would impact host-switched *Varroa* populations. Alternatively, *de novo* mutation rates could contribute to population genetic variation, but our direct estimation of mutation rate did not show exceptionally high mutation rates (>8 × 10^{-10} per bp per generation), compared to other arthropods (31). Consequently, immigration may be the primary means of introducing new alleles into populations on the novel host during early stages of the host switch.

While host switch events by specialized parasites were once assumed to be rare, this assumption is increasingly challenged (32, 33). Ecological fitting theory suggests that shifts could readily occur in species with a pre-existing ability to use novel hosts (34). Host switches in *Varroa* appear to fit this model. While the western honey bee has been introduced throughout the *Varroa* native range, mitochondrial data, which are geographically informative (Figure 1), suggest that switches occurred only in Korea, Japan and the Philippines (*V. destructor*) and Papua New Guinea (*V. jacobsoni*) (3, 10, 11). *A. cerana* subspecies are strongly differentiated geographically (35–39), and *Varroa* mitotypes mirror host biogeography and subspecies

distribution (9, 11, 40). This suggests that *Varroa* populations may vary in traits, such as host specificity, as a result of their previous coevolutionary interaction with local *A. cerana* subspecies. This may have allowed some populations to switch, while others were unable, in spite of available novel hosts (41). However, the dynamic nature of the interaction does not preclude additional switches in the future.

One potential limitation of our data set is its sparse geographic sampling, given the size of the native range. However, the numerous markers available for each individual allow accurate demographic modeling (42, 43), even with sample sizes as small as two individuals (44, 45). While larger samples should increase the confidence in model selection and parameter estimates, the sampling design followed established recommendations (46). We captured most of the described *Varroa* mitochondrial "haplogroups" and even discovered new ones, (2, 10, 12, 47), with the notable exception of the host-switched Japanese (J1) haplotype. In the future, incorporating mites from this source population could help estimate its genetic contribution to host switch success, providing that the J1 lineage was not completely displaced (48, 49). Despite these limitations, we reconstructed known aspects of *Varroa* host switch demography, such as the times of the switches. However, these times need to be regarded as approximate, since generation times may fluctuate due to a variety of factors, such as seasonality or brood availability. We relied on a previous approximation of ten generations per year (50).

Our findings highlight the dynamic and ongoing nature of host switches in the native range, and the need to better understand native mite populations. Mitochondrial DNA has already detected the likely presence of several host switches, such as those in Philippines in 2015 (11) and the presence of new haplogroups in Eastern Europe (51), indicating that populations of *Varroa* are continuously testing *A. mellifera* as a new host, and may have spread without our awareness. Additionally, while this work provides insight into the initial host switch, the demographics of the subsequent worldwide pandemic remain largely unknown, for instance the critical population size necessary to establish a regional infestation. Future research on global demographic parameters such as genetic diversity and gene flow is crucial (52). Since *V. jacobsoni* presents striking parallels to *V. destructor*, this knowledge could be applied to forecast and prevent its spread.

Materials and Methods

Mite sampling on original and novel honey bee hosts

V. destructor specimens were collected between 1996 and 2003 as part of the taxonomic revision and diversity surveys carried by Anderson and Trueman (2) and Navajas *et* al. (10) (Table 1). *V. jacobsoni* specimens were collected between 2008 and 2015 as part of a field survey for reproduction and host shift by Roberts et al (3). All individuals were mature sclerotized female mites collected from different colonies and were preserved in individual Eppendorf tubes labeled with collection information, containing 70% ethanol, and kept at -20°C at the CSIRO Canberra, Australia. We obtained 25 *Varroa* mites from eastern honey bee, *A. cerana,* colonies and 18 from western honey bee, *A. mellifera*, colonies (Table S1). Sampling coverage encompassed as many geographic regions as available from the *Varroa* native range, for which both original *A. cerana* and novel *A. mellifera* hosts occur in sympatry (Figure 1A). Exact geo-coordinates were not always available for samples collected in China, South Korea, Viet Nam, Thailand and Indonesia before 2008, and were approximated by the locality provided, and guided by the earliest *Varroa* survey maps (2, 10, 47, 53).

DNA extraction and whole-genome sequencing

Mites were surface sterilized by cleaning them in absolute ethanol using a sterile brush to remove any external debris, and then gently shaken in a 2.0 mL Eppendorf filled with absolute ethanol. Each mite was then dried for 10 sec on a sterile paper towel before being placed in a 1.5 mL Eppendorf tube in liquid nitrogen. Genomic DNA was extracted from each mite by crushing the whole body, using a sterile pestle to obtain a fine powder and processed with a QIAamp DNA Micro Kit (© Qiagen) following the manufacturer's instructions. Final elution volume was 15 µL. Total dsDNA was measured using a Qubit™ 4 Fluorometer with an Invitrogen dsDNA HS Assay Kit.

For population genomic samples (Table S2), short-inserts of 150-bp paired-end libraries were prepared for each individual using a Nextera XT DNA Library Preparation Kit (Illumina ®). Size-selection and cleanup were accomplished using CA-magnetic beads (Dynabeads® MyOne Carboxylic Acid, Invitrogen), and 11-11.5% PEG 6000 (Sigma-Aldrich © LLC). Library quality and size were assessed using a Bioanalyzer High Sensitivity DNA kit (Agilent). All libraries were pooled and sequenced on six lanes of a HiSeq 4000 (Illumina ®) at the OIST Sequencing Center.

Alignment and Variant calling

Detailed command lines used for each analysis step are implemented on our Snakemake script (54) available on https://github.com/MaevaTecher/varroa-host-jump. Briefly, we assessed demultiplexed fastq read quality using FastQC (55) and outliers were quickly visualized using Mashtree (56). We mapped reads to the *V. destructor* reference genome on NCBI [GCF_002443255.1] (16) separately from the complete mitogenome [NC_004454.2] (57) using Bowtie2 v2.6 (58) and NextGenMap v0.5.0 in parallel (59). Reads were sorted and duplicates were removed using SAMtools (60), and subsampled to a maximum coverage of 500 using VariantBam (61). Then, we performed variant calling using FreeBayes v1.1.0 (62) with the following parameters: minimum mapping and base quality was 5, evaluate only the best four SNP alleles, and populations are *Varroa* species. Variants were subsequently filtered in the following order: allele quality score >20, at least 90% of samples had data for the variant, and a coverage cut-off depending on the read depth was applied using vcflib v1.0.0-rc1 (63). Then using VCFtools v0.1.12b (64) we removed variants found in genome masked regions (highly repetitive, low complexity or gaps), indels, sites with more than two alleles and with a minor allele frequency lower than 0.05. Mapping and variant calls of the mitochondrial genome followed the same process except that haploidy was indicated in FreeBayes parameters.

Species identification using standard mitochondrial barcoding

To accurately determine *Varroa* species and haplogroup identity, we used the following standard mitochondrial markers: 1) a 458-bp fragment from the *COX1* gene (position in NC_004454.2: 1817-2275 bp) (2, 17, 53) and 2) a concatenated 2,696-bp fragment from the four genes *COX1* (NC_004454.2: 1,187-2,275), *ATP6* (NC_004454.2: 4,008-4,343), *COX3* (NC_004454.2: 4,348-4,722), and *CYTB* (NC_004454.2: 9,847-10,746) (10). Following variant call and filtering on the mitochondrial genome, a consensus fasta sequence was generated for each individual using the option vcf2fasta by vcflib. Regions of interest were extracted and screened on Geneious Prime® 2019.2.3.

To ensure that filtering did not remove existing and known variants, we also directly sequenced the *COX1* gene. We used the primers 10kbCOIF1 (5'- CTT GTA ATC ATA AGG ATA TTG GAAC -3') and 6,5KbCOIR (5'- AAT ACC AGT GGG AAC CGC - 3') (10). PCR reactions were carried out in 25µL containing 5 µL of 5X Phusion® HF buffer; 0.5 µL of dNTP mix (10mM); 0.25 µL of Phusion® High-Fidelity DNA Polymerase (NEB); 1.25 µL of each oligo primer 10kbCOIF1and 6,5KbCOIR (10mM);

1 µL of template DNA (0.5 ng/µL) and Milli-Q water. Samples were denatured at 98°C for 30 sec, and then PCR was performed for 35 cycles of 10 sec denaturation at 98°C, 15 sec of annealing at 59°C and 15 sec of extension at 72°C with a 5 min final elongation at 72°C. DNA amplification success was visualized by loading 3 µL of PCR product with 3 µL of loading dye on 1% agarose gel (110V for 20 min). PCR products were then cleaned-up using Dynabeads® MyOne Carboxylic Acid, CA-beads (Invitrogen) and 19% Polyethylene glycol PEG. Directly, fragments were sequenced in the two directions using the BigDye™ Terminator v3.1 Cycle Sequencing Kit (Thermo Fisher) in a capillary sequencer Applied Biosystems 3730xl DNA Analyzer (Thermo Fisher). FASTA sequences generated by mitogenome mapping and Sanger sequencing were then aligned and checked for discordant nucleotide variation.

Phylogenetic relationships with 60 unique *COX1* reference sequences (4), were assessed after alignment using ClustalW and manual check. Given the existing variability in sequence length from the reference sequences region, a neighbor-joining tree was built using the Tamura-Nei genetic distance model with *V. rindereri* [AF107261] (2) as an outgroup and branch support was estimated with 1,000 bootstraps. Regarding the *COX1-ATP6-COX3-CYTB* concatenated region, sequences were aligned with 22 reference concatenated sequences for *V. destructor* only (10, 51, 65).. A haplotype median-joining network was drawn using PopART (66) after following software suggestions to remove sequences with more than 5% overall site gap (Serbia1 and Peshter1).

Inference of population structure and genetic diversity

Genetic diversity and structure were explored within and among species considering different population unit levels: species, species-honey bee host and species-honey bee host-mtDNA lineages. Nucleotide diversity (π), population genetic differentiation (F_{ST}) and the number of nucleotide substitutions (D_{xy}) between each population pair were computed using a sliding window (50kb window, sliding every 20kb and containing a minimum of 100 SNPs). This method was implemented using python scripts transforming and required the allelic primitive vcf including monomorphic sites for D_{xy} computation (before MAF filtering) github.com/simonhmartin/genomics_general) (67). Mean pairwise F_{ST} per site and considering several population levels were estimated using VCFtools. As the threshold for minor allele frequency (MAF) filtering can affect the observed population structure (68), genetic indices and differentiation analysis were done for both pre-filter and filter-LD pruned MAF datasets.

Subsequently, genetic structure was also assessed using a principal component analysis (PCA) performed with the R package vcfR (69) on the variant dataset generated after minor allele frequency filtering. Then, PLINK v1.90b3 (70) was used to conduct PCA on a trimmed subset that minimized linkage disequilibrium (LD) between markers by retaining only one pair of variants when LD was greater than 0.5, considering a window of 50 SNPs with a shift of 5 SNPs. Additionally, population cluster analysis was led using genotype likelihoods extracted from the vcf file with VCFtools and the BEAGLE format to input into NGSadmix (71). The probabilistic framework considering K = 2 to 20 genetic clusters was evaluated using 5 replicates. The number of genetic clusters among and within species was determined following guidelines in (72, 73).

Checking for evidence of migration between hosts

We cross-checked honey bee host identities reported during sampling with host read identity to detect potentially drifting mites. As mites feed on honey bees during the phoretic phase (74) and maintain a consistent feeding regimen by consuming ~1 µL of host fluid (digested fat-body and haemolymph) per day (75), we assumed that host DNA would be retrieved from crushed mite

tissues. Mitochondrial DNA was targeted, as it is more abundant than nuclear DNA. We mapped raw fastq reads on honey bee host mitochondrial reference genomes using [NC_001566.1] for *A. mellifera ligustica* (76) and [NC_014295] for *A. cerana (77)*. The number of reads mapped to either one of these honey bee host reference genomes was counted and compared to sampled host identities.

Mutation rate estimation from mother-son pairs

We attempted to estimate mutation rates for *V. destructor* by genotyping mothers and their haploid sons. These were collected from single capped cells containing *A. mellifera* red-eye drone pupae. Each family was selected under the following conditions: 1) each cell was only infected by a single mother-foundress mite, 2) a family at this stage should have contained three to four offspring, including one haploid son and two to three diploid daughters, 3) the son was at the deutonymph or adult phase to avoid misidentification with protonymph sisters. Thus, three families were collected and preserved in absolute ethanol in February 2018 (#1, #2, and #7), and three more in May 2018 (#15, #17, and #19) from the same colony at the OIST Ecology and Evolution experimental apiary.

Libraries were prepared using an NEBNext® Ultra™ II FS DNA Library Prep Kit (New England Biolabs, Inc) with a fragmentation size incubation time of 15 min at 37°C, and three PCR cycles. Libraries were cleaned up using CA-magnetic beads (Dynabeads® MyOne Carboxylic Acid, Invitrogen), and 17% PEG 6000 (Sigma-Aldrich © LLC). Libraries were pooled and sequenced on two lanes of HiSeq 4000 (Illumina ®) at the OIST sequencing center.

Mutation rate was estimated from GATK variant calls using DeNovoGear's dng-call algorithm, which can model *Varroa*'s haplodiploidy sex-determination system. Single-nucleotide mutations were called on the 7 largest contigs in the *Varroa* genome, avoiding sites that were within 100 bp of an indel. Mutation calls were de-duplicated, retaining only sites that were biallelic and not part of a mutation cluster (with 100 bp of another call or part of a long-run of calls from the same sample). After deduplication, the remaining calls were filtered such that P(denovo| data) was high (DNP >= 0.75) and the fit to the model was good (LLS >= -3). DNP (de-novo probability) and LLS (log-likelihood scaled) are both per-site statistics generated by dng-call. The denominator for the mutation rate analysis was estimated by simulating mutations at 1,000 locations in each son and calculating what fraction of these simulated mutations was recovered by our pipeline. A VCF was generated for these locations, and mutations were simulated by changing the son's haplotype to one of the other three bases. ALT and AD fields were updated as needed.

Divergence and demographic analyses

We inferred the demographic history of both *V. destructor* and *V. jacobsoni* mites on *A. cerana* and *A. mellifera* with the coalescent simulator fastsimcoal 2.6 using site frequency spectrum (SFS) dataset (78). Four subsets of ~1,050, 10,500, 35,000 and 105,000 SNPs were randomly selected from the LD-pruned dataset using BCFtools for each *Varroa* species and included only *loci* with minor allele frequencies over 0.05 for the selected individuals. For *V. destructor* datasets, two host populations were considered with *A. mellifera* (n = 8) and *A. cerana* (n = 9) with the exclusion of samples exhibiting misassigned host signals. For *V. jacobsoni*, only mites occurring in Papua New Guinea were analyzed by also removing misassigned host individuals (n = 8 on *A. mellifera*, n = 4 on *A. cerana*). The joint folded SFS for each host population pair was projected using the easySFS python scripts (github.com/isaacovercast/easySFS) and input directly into fastsimcoal 2. All subsequent demographic inference computations were based on the estimated result of mutation rate $\mu = 1.25 \times 10^{-8}$ / bp/generation.

Six scenarios were designed in light of known host switches and existing artificial sympatry between original and novel honey bee hosts in Asia and Oceania (Figure 3). In all scenarios, the estimated effective population size (*Ne*) for *Varroa* mites parasitizing *A. cerana* NVAC (in haploid individuals) was considered stable before, during, and after the host switch bottleneck, occurring at a time TBOT (in generations). In scenario 1, size of the newly founded population of *Varroa* mites on the new host, *A. mellifera* NVAM (in haploid individuals), was considered the same from TBOT to the present day. In scenario 2, the *Varroa* population on *A. mellifera* was forced to start from a single mite at TBOT, and expanded to the present-day population size NVAM with a growth rate GAM. Scenario 3 followed the same scheme as scenario 2, but the initial population size NJUMP (in haploid individual) colonizing *A. mellifera* was freely estimated. Scenarios 4, 5, and 6, mirrored the first three scenarios, but considered bilateral migration among host populations possible (MAMtoAC/MACtoAM = proportion of haploid individuals migrating from one population to another). Search ranges for each parameter imputed were wide with uniform distribution (although there is no real upper range limit in fastsimcoal 2), and only the TBOT rule was changed from one species to another.

For each model, fastsimcoal 2 computed the expected SFS and the parameters from the highest likelihood obtained after 100 loops (e.g., ECM cycles, --numloops 100) with one million coalescent simulations per loop (--numsims 1,000,000), and with a 0.001 minimum relative difference in parameter values estimated by the maximum composite likelihood from the SFS (--maxlhood 0.001). A minimum of 20 loops were performed, for which the likelihood was computed from both monomorphic and polymorphic sites (--minnumloops 20) and SFS entries with fewer than 10 SNPs were pooled together (--minSFSCount 10). This process was replicated 100 times for each model and the set of estimated parameters with the highest likelihood was retained as the best point estimate.

For each model, we calculated the AIC to identify the best fitting demographic scenario (78). In parallel, we compared the likelihood distribution for each model to evaluate whether other models than the one with the best AIC could also fit the observed data (79). 95% confidence intervals were obtained from the maximum likelihood parameters estimated using parametric bootstraps. The best estimate parameters were used to generate 100 pseudo-observed SFS (for a similar number of polymorphic SNPs) and estimating the parameters following the same process as stated before. The mean and 95 percentile confidence intervals were computed using the 'boot' R package (80). Finally, the fit of the best expected SFS was visualized against the observed using SFStools R scripts (github.com/marqueda/SFS-scripts).

Acknowledgments

MAT's research was supported by a postdoctoral fellowship from the Japan Society for Promotion of Science (JSPS) (P19723), Kakenhi Grant-in-Aid, (19F19723) and the Okinawa Institute of Science and Technology (OIST). ASM was supported by a Future Fellowship from the Australian Research Council (FT160100178) and a Kakenhi Grant-in-Aid for Scientific Research from the JSPS (18H02216). JMKR was supported by the Australian Centre for International Agricultural Research (ACIAR) and the Australian Department of Agriculture, Water and Environment (DAWE). RAC was supported by National Institutes of Health (R01-HG007178). We are grateful to Jo Si Lay Tan and Lijun Qiu for their advice and guidance in developing the wet lab workflow for *Varroa* mite sequencing. We also would like to thank the OIST Sequencing Center for assisting us in the sample sequencing. We wish to thank Steven D. Aird, technical editor, for reviewing and improving our manuscript.

Data accessibility

Raw FASTQ files for all individuals sequenced were submitted to the DNA Databank of Japan and uploaded to GenBank under the bioproject PRJDB9195 with the run accessions DRR209082-DRR209125 and DRR212369-DRR212380. The pipeline developed using Snakemake, Rmarkdown for data exploration, demographic inferences input files and FASTA sequences alignment (mtDNA) are made available for reproducibility on github.com/MaevaTecher/varroa-host-jump. Variant calling files, genome indexing files and input lists are readily available at DRYAD.

Author Contributions

MAT and ASM designed research, analyzed population genetics data and wrote the manuscript. MAT processed the samples in the wet lab until library preparation and ran the demographic inferences. JMKR collected and provided samples from the mite CSIRO collection, and helped in the data interpretation. RAC analyzed data, wrote the manuscript, provided reproducible online resources and data for the estimation of mutation rate.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. M. J. Hatcher, J. T. A. Dick, A. M. Dunn, Disease emergence and invasions. *Funct. Ecol.* **26**, 1275–1287 (2012).

2. D. L. Anderson, J. W. Trueman, Varroa jacobsoni (Acari: Varroidae) is more than one species. *Exp. Appl. Acarol.* **24**, 165–189 (2000).

3. J. M. K. Roberts, D. L. Anderson, W. T. Tay, Multiple host shifts by the emerging honeybee parasite, Varroa jacobsoni. *Mol. Ecol.* **24**, 2379–2391 (2015).

4. K. Traynor, *et al.*, Varroa destructor: A Complex Parasite, Crippling Honeybees Worldwide. *BIOLOGY* (2020) https:/doi.org/10.20944/preprints202002.0374.v1.

5. A. Noël, Y. Le Conte, F. Mondet, Varroa destructor: how does it harm Apis mellifera honey bees and what can be done about it? *Emerg Top Life Sci* (2020) https:/doi.org/10.1042/ETLS20190125.

6. E. Crane, Beekeeping round the World. *Bee World* **49**, 113–114 (1968).

7. P. Chantawannakul, L. I. de Guzman, J. Li, G. R. Williams, Parasites, pathogens, and pests of honeybees in Asia. *Apidologie* **47**, 301–324 (2016).

8. M. Solignac, *et al.*, The invasive Korea and Japan types of Varroa destructor, ectoparasitic mites of the Western honeybee (Apis mellifera), are two partly isolated clones. *Proc. Biol. Sci.* **272**, 411–419 (2005).

9. N. Warrit, D. R. Smith, C. Lekprayoon, Genetic subpopulations of Varroa mites and their Apis cerana hosts in Thailand. *Apidologie* **37**, 19–30 (2006).

10. M. Navajas, *et al.*, New Asian types of Varroa destructor: a potential new threat for world apiculture. *Apidologie* **41**, 181–193 (2010).

11. A. L. Beaurepaire, *et al.*, Host Specificity in the Honeybee Parasitic Mite, Varroa spp. in Apis mellifera and Apis cerana. *PLoS One* **10**, e0135103 (2015).

12. V. Dietemann, *et al.*, Population genetics of ectoparasitic mites Varroa spp. in Eastern and Western honey bees. *Parasitology* **146**, 1429–1439 (2019).

13. J. González-Cabrera, *et al.*, Novel Mutations in the Voltage-Gated Sodium Channel of Pyrethroid-Resistant Varroa destructor Populations from the Southeastern USA. *PLoS One* **11**, e0155332 (2016).

14. W. E. Fry, Phytophthora infestans: New Tools (and Old Ones) Lead to New Understanding and Precision Management. *Annu. Rev. Phytopathol.* **54**, 529–547 (2016).

15. M. McMullan, *et al.*, The ash dieback invasion of Europe was founded by two genetically divergent individuals. *Nat Ecol Evol* **2**, 1000–1008 (2018).

16. M. A. Techer, *et al.*, Divergent evolutionary trajectories following speciation in two ectoparasitic honey bee mites. *Commun Biol* **2**, 357 (2019).

17. V. Dietemann, *et al.*, Standard methods for varroa research. *J. Apic. Res.* **52**, 1–54 (2013).

18. B. Gajić, *et al.*, Haplotype identification and detection of mitochondrial DNA heteroplasmy in Varroa destructor mites using ARMS and PCR-RFLP methods. *Exp. Appl. Acarol.* **70**, 287–297 (2016).

19. R. Poulin, B. R. Krasnov, D. Mouillot, D. W. Thieltges, The comparative ecology and biogeography of parasites. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **366**, 2379–2390 (2011).

20. R. M. Penczykowski, A.-L. Laine, B. Koskella, Understanding the ecology and evolution of host-parasite interactions across scales. *Evol. Appl.* **9**, 37–52 (2016).

21. J. P. Sexton, J. Montiel, J. E. Shay, M. R. Stephens, R. A. Slatyer, Evolution of Ecological Niche Breadth. *Annu. Rev. Ecol. Evol. Syst.* **48**, 183–206 (2017).

22. B. A. McDonald, C. Linde, Pathogen population genetics, evolutionary potential, and durable resistance. *Annu. Rev. Phytopathol.* **40**, 349–379 (2002).

23. M.-A. Auger-Rozenberg, *et al.*, Inferences on population history of a seed chalcid wasp: invasion success despite a severe founder effect from an unexpected source population. *Mol. Ecol.* **21**, 6086–6103 (2012).

24. J. W. Demastes, D. J. Hafner, M. S. Hafner, J. E. Light, T. A. Spradling, Loss of genetic diversity, recovery and allele surfing in a colonizing parasite, Geomydoecus aurei. *Mol. Ecol.* **28**, 703–720 (2019).

25. A. L. Beaurepaire, *et al.*, Population genetics of ectoparasitic mites suggest arms race with honeybee hosts. *Sci. Rep.* **9**, 11355 (2019).

26. C. Pinho, J. Hey, Divergence with Gene Flow: Models and Data. *Annu. Rev. Ecol. Evol. Syst.* **41**, 215–230 (2010).

27. J. L. Feder, S. P. Egan, P. Nosil, The genomics of speciation-with-gene-flow. *Trends Genet.* **28**, 342–350 (2012).

28. M. Tomasini, S. Peischl, When does gene flow facilitate evolutionary rescue? *Evolution* (2020) https:/doi.org/10.1111/evo.14038.

29. T. Lenormand, Gene flow and the limits to natural selection. *Trends Ecol. Evol.* **17**, 183–189 (2002).

30. A. R. Whiteley, S. W. Fitzpatrick, W. C. Funk, D. A. Tallmon, Genetic rescue to the rescue. *Trends Ecol. Evol.* **30**, 42–49 (2015).

31. H. Liu, *et al.*, Direct Determination of the Mutation Rate in the Bumblebee Reveals Evidence for Weak Recombination-Associated Mutation and an Approximate Rate Constancy in Insects. *Mol. Biol. Evol.* **34**, 119–130 (2017).

32. T. Rigaud, M.-J. Perrot-Minnot, M. J. F. Brown, Parasite and host assemblages: embracing the reality will improve our knowledge of parasite transmission and virulence. *Proc. Biol. Sci.* **277**, 3693–3702 (2010).

33. S. Nylin, *et al.*, Embracing Colonizations: A New Paradigm for Species Association Dynamics. *Trends Ecol. Evol.* **33**, 4–14 (2018).

34. S. B. L. Araujo, *et al.*, Understanding Host-Switching by Ecological Fitting. *PLoS One* **10**, e0139225 (2015).

35. S. E. Radloff, *et al.*, Population structure and classification of Apis cerana. *Apidologie* **41**, 589–601 (2010).

36. K. Tan, Y. Qu, Z. Wang, Z. Liu, M. S. Engel, Haplotype diversity and genetic similarity among populations of the Eastern honey bee from Himalaya-Southwest China and Nepal (Hymenoptera: Apidae). *Apidologie* **47**, 197–205 (2016).

37. C. Chen, *et al.*, Population genomics provide insights into the evolution and adaptation of the eastern honey bee (Apis cerana). *Mol. Biol. Evol.* (2018) https:/doi.org/10.1093/molbev/msy130.

38. L. Yancan, C. Tianle, F. Yunhan, L. Delong, W. Guizhi, Population genomics and morphological features underlying the adaptive evolution of the eastern honey bee (Apis cerana). *BMC Genomics* **20**, 869 (2019).

39. R. A. Ilyasov, *et al.*, Phylogenetic Relationships of Russian Far-East Apis cerana with Other North Asian Populations. *Journal of Apicultural Science* **63**, 289–314 (2019).

40. O. Rueppell, A. M. Hayes, N. Warrit, D. R. Smith, Population structure of Apis cerana in Thailand reflects biogeography and current gene flow rather than Varroa mite association. *Insectes Soc.* **58**, 445–452 (2011).

41. W. Li, C. Wang, Z. Y. Huang, Y. Chen, R. Han, Reproduction of Distinct Varroa destructor Genotypes on Honey Bee Worker Brood. *Insects* **10**, 372 (2019).

42. J. E. Pool, I. Hellmann, J. D. Jensen, R. Nielsen, Population genetic inference from genomic sequence variation. *Genome Res.* **20**, 291–300 (2010).

43. A. C. Beichman, E. Huerta-Sanchez, K. E. Lohmueller, Using Genomic Data to Infer Historic Population Dynamics of Nonmodel Organisms. *Annu. Rev. Ecol. Evol. Syst.* **49**, 433–456 (2018).

44. A. G. Nazareno, J. B. Bemmels, C. W. Dick, L. G. Lohmann, Minimum sample sizes for population genomics: an empirical study from an Amazonian plant species. *Mol. Ecol. Resour.* **17**, 1136–1147 (2017).

45. J. Salmona, R. Heller, M. Lascoux, A. Shafer, "Inferring Demographic History Using Genomic Data: Concepts, Approaches and Applications" in *Population Genomics*, Population Genomics., O. P. Rajora, Ed. (Springer International Publishing, 2019), pp. 511–537.

46. J. D. Robinson, A. J. Coffman, M. J. Hickerson, R. N. Gutenkunst, Sampling strategies for frequency spectrum-based population genomic inference. *BMC Evol. Biol.* **14**, 254 (2014).

47. T. Zhou, *et al.*, Identification of Varroa mites (Acari: Varroidae) infesting Apis cerana and Apis mellifera in China. *Apidologie* **35**, 645–654 (2004).

48. L. I. de Guzman, T. E. Rinderer, J. Anthony Stelzer, Occurrence of two genotypes of Varroa jacobsoni Oud. in North America. *Apidologie* **30**, 31–36 (1999).

49. M. H. Ogihara, M. Yoshiyama, N. Morimoto, K. Kimura, Dominant honeybee colony infestation by Varroa destructor (Acari: Varroidae) K haplotype in Japan. *Appl. Entomol. Zool.* **55**, 189–197 (2020).

50. J. M. Cornuet, M. A. Beaumont, A. Estoup, M. Solignac, Inference on microsatellite mutation processes in the invasive mite, Varroa destructor, using reversible jump Markov chain Monte Carlo. *Theor. Popul. Biol.* **69**, 129–144 (2006).

51. B. Gajic, *et al.*, Variability of the honey bee mite Varroa destructor in Serbia, based on mtDNA analysis. *Exp. Appl. Acarol.* **61**, 97–105 (2013).

52. J. D. Evans, S. C. Cook, Genetics and physiology of Varroa mites. *Curr Opin Insect Sci* **26**, 130–135 (2018).

53. D. L. Anderson, S. Fuchs, Two genetically distinct populations of Varroa jacobsoni with contrasting reproductive abilities on Apis mellifera. *J. Apic. Res.* **37**, 69–78 (1998).

54. J. Köster, S. Rahmann, Snakemake--a scalable bioinformatics workflow engine. *Bioinformatics* **28**, 2520–2522 (2012).

55. S. Andrews, Others, FastQC: a quality control tool for high throughput sequence data (2010).

56. L. Katz, *et al.*, Mashtree: a rapid comparison of whole genome sequence files. *JOSS* **4**, 1762 (2019).

57. M. Navajas, Y. Le Conte, M. Solignac, S. Cros-Arteil, J.-M. Cornuet, The complete sequence of the mitochondrial genome of the honeybee ectoparasite mite Varroa destructor (Acari: Mesostigmata). *Mol. Biol. Evol.* **19**, 2313–2317 (2002).

58. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

59. F. J. Sedlazeck, P. Rescheneder, A. von Haeseler, NextGenMap: fast and accurate read mapping in highly polymorphic genomes. *Bioinformatics* **29**, 2790–2791 (2013).

60. H. Li, *et al.*, The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

61. J. Wala, C.-Z. Zhang, M. Meyerson, R. Beroukhim, VariantBam: filtering and profiling of next-generational sequencing data using region-specific rules. *Bioinformatics* **32**, 2029–2031 (2016).

62. E. Garrison, G. Marth, Haplotype-based variant detection from short-read sequencing. *arXiv [q-bio.GN]* (2012).

63. E. Garrison, Vcflib: A C++ library for parsing and manipulating VCF files. *GitHub https://github.com/ekg/vcflib* (2012).

64. P. Danecek, *et al.*, The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).

65. T. Elbeaino, *et al.*, Occurrence of Deformed wing virus, Chronic bee paralysis virus and mtDNA variants in haplotype K of Varroa destructor mites in Syrian apiaries. *Exp. Appl. Acarol.* **69**, 11–19 (2016).

66. J. W. Leigh, D. Bryant - Methods in Ecology and Evolution, 2015, popart: full-feature software for haplotype network construction. *Wiley Online Library* (2015).

67. S. H. Martin, J. W. Davey, C. D. Jiggins, Evaluating the use of ABBA-BABA statistics to locate introgressed loci. *Mol. Biol. Evol.* **32**, 244–257 (2015).

68. E. Linck, C. J. Battey, Minor allele frequency thresholds strongly affect population structure inference with genomic datasets. *Mol. Ecol. Resour.* (2019).

69. B. J. Knaus, N. J. Grünwald, vcfr: a package to manipulate and visualize variant call format data in R. *Mol. Ecol. Resour.* **17**, 44–53 (2017).

70. S. Purcell, *et al.*, PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).

71. L. Skotte, T. S. Korneliussen, A. Albrechtsen, Estimating individual admixture proportions from next generation sequencing data. *Genetics* **195**, 693–702 (2013).

72. P. G. Meirmans, Seven common mistakes in population genetics and how to avoid them. *Mol. Ecol.* **24**, 3223–3231 (2015).

73. D. J. Lawson, L. van Dorp, D. Falush, A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. *Nat. Commun.* **9**, 3258 (2018).

74. S. D. Ramsey, *et al.*, Varroa destructor feeds primarily on honey bee fat body tissue and not hemolymph. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 1792–1801 (2019).

75. F. Posada-Florez, *et al.*, Insights into the metabolism and behaviour of Varroa destructor mites from analysis of their waste excretions. *Parasitology* **146**, 527–532 (2019).

76. R. H. Crozier, Y. C. Crozier, The mitochondrial genome of the honeybee Apis mellifera: complete sequence and genome organization. *Genetics* **133**, 97–117 (1993).

77. H.-W. Tan, *et al.*, The complete mitochondrial genome of the Asiatic cavity-nesting honeybee Apis cerana (Hymenoptera: Apidae). *PLoS One* **6**, e23008 (2011).

78. L. Excoffier, I. Dupanloup, E. Huerta-Sánchez, V. C. Sousa, M. Foll, Robust demographic inference from genomic and SNP data. *PLoS Genet.* **9**, e1003905 (2013).

79. J. I. Meier, *et al.*, Demographic modelling with whole-genome data reveals parallel origin of similar Pundamilia cichlid species after hybridization. *Mol. Ecol.* **26**, 123–141 (2017).

80. A. Canty, B. Ripley, Package "boot" (2019).