

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. Roberts², Reed A Cartwright^{3,4}, and Alexander S. Mikheyev^{1,5*}

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

10 † Current address: Texas A&M University, Department of Entomology, Biological Control Facility, College Station, Texas
77840, USA

Maeva A. Techer 0000-0001-5417-5103

maeva.angelique.techer@gmail.com

15 John M. K. Roberts 0000-0001-9739-5595

John.Roberts@csiro.au

Reed A. Cartwright 0000-0002-0837-9380

cartwright@asu.edu

Alexander S. Mikheyev 0000-0003-4369-1019

20 alexander.mikheyev@anu.edu.au

Corresponding Authors *

Maeva A. Techer maeva.techer@ag.tamu.edu

Alexander S. Mikheyev alexander.mikheyev@anu.edu.au

Abstract

25 **Background:** Host switching allows parasites to expand their niches. However, successful switching may require suites of adaptations and also 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
30 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*).

Results: 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
35 worldwide collapses of honey bee populations. Using whole-genome sequencing on 63 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 multiple previously undiscovered mitochondrial lineages on the novel host, along with the genetic equivalent of tens of individuals that were involved in the initial host switch. Despite being greatly reduced, some gene flow remains between mites adapted to different hosts.

40 **Conclusion:** Our findings suggest that while reproductive isolation may facilitate the fixation of traits beneficial for exploitation of the new host, ongoing genetic exchange may allow genetic amelioration of inbreeding effects.

Keywords

Host switch, *Varroa spp.*, honey bees, whole-genome sequencing, population genomics, founder size, sympatry.

Background

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 and behaviour 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 [1]. Globalization has eased these barriers, sometimes giving rise to pandemics [2]. 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* recently followed by *Varroa jacobsoni*, which acquired the western honey bee (*Apis mellifera*) as a new host, ~70 and ~12 years ago, respectively [3,4]. *V. destructor*, in particular, spread worldwide, causing extensive honey bee population collapses, whereas *V. jacobsoni* has so far remained in Oceania [5,6]. Both mites were originally found on the sister species, *Apis cerana*, and came into contact with growing populations of *A. mellifera*, which were brought in for purposes of beekeeping [7,8]. 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 relatively short mitochondrial markers, which revealed that in both species populations on the new host were strongly differentiated and genetically depauperate [4,9–13]. While *V. destructor*, in particular, was described as “quasi-clonal” and highly inbred [9], 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 synthetic acaricides How does

a bottlenecked species with high inbreeding 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]

75 This is also true of many invasive animals who overcome genetic bottlenecks with repeated introductions or other mechanisms such as multiple mating [16–19], making the case of *V. destructor* that much more puzzling.

To answer this question and to gain broader insight into how host switches happen, we sequenced nuclear and mitochondrial 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
80 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 spp.* succeed.

Results

Cryptic diversity in mitogenomes suggests multiple foundresses at the origins of the host switch

85 We examined the genome-wide variation and divergence among 31 *V. destructor* and 32 *V. jacobsoni* females from their original and novel hosts (Table 1). *Varroa* species identity was confirmed by extracting and aligning mitogenomes together with known reference sequences of the mtDNA COX1 458-bp standard marker (Table S1 and Figure S1). A sole mismatch between taxonomic identification and mtDNA barcoding was detected in a *V. jacobsoni* specimen (THV003_1), which was collected in Thailand on *A. mellifera*, though it is not known whether
90 the mite could reproduce on this honey bee species.

Using the set of 2,091 SNPs detected across the mitogenome, we reconstructed phylogenetic relationships among *Varroa* lineages in each species (Figure 1). While exhibiting more genomic variability than previously reported, the mitogenome networks were consistent with the standard classification of *Varroa spp.* geographic lineages. We found that the consensus sequences of both species diverged 3.6%. Divergence within *V. destructor* samples was very low (0.5% with 85 SNPs), but moderate in *V. jacobsoni* with up to 1.3% (216 SNPs).
95

Table 1: *Varroa spp.* specimens used for population genomics were collected from their native range across 11 countries, both on their original or new honey bee host. Newly reported mtDNA lineages are indicated in bold.

Species	Host	Country	Date	N	mtDNA lineages as defined by standard <i>COX1</i>
<i>V. destructor</i>	Original host <i>A. cerana</i>	Japan	1994	2	VD Japan J1
		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	2	VD Viet Nam V1 VD Korea K1
		Myanmar	2003	1	VD Viet Nam V1
		Nepal	2000	1	VD Nepal 2
<i>V. destructor</i>	New host <i>A. mellifera</i>	Japan	1996	1	VD Korea K1
		South Korea	1996	3	VD Korea K1
		China	2001	5	VD Korea K1
		Viet Nam	1996, 1998	6	VD Korea K1
		Myanmar	2003	2	VD Korea K1
<i>V. jacobsoni</i>	Original host <i>A. cerana</i>	India	2002	2	VJ Bengaluru 1
		Thailand	2003	1	VJ Laos L1
		Viet Nam	2003-2004	5	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
<i>V. jacobsoni</i>	Original host <i>A. nigrocincta</i>	Indonesia	1996	3	VJ Java 1
<i>V. jacobsoni</i>	New host <i>A. mellifera</i>	Thailand	1994	1	VJ North Thailand 4
		Papua New Guinea	2008, 2014	9	VJ PNG 1

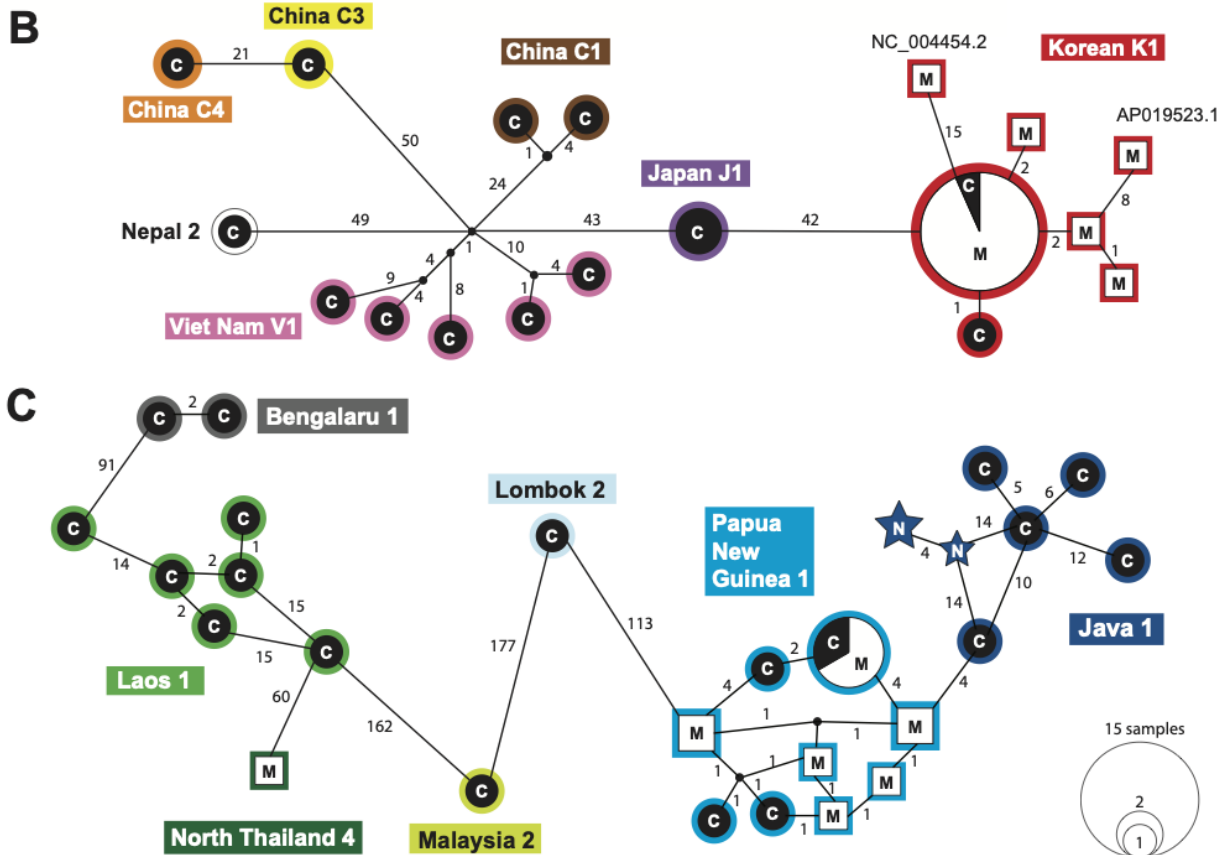
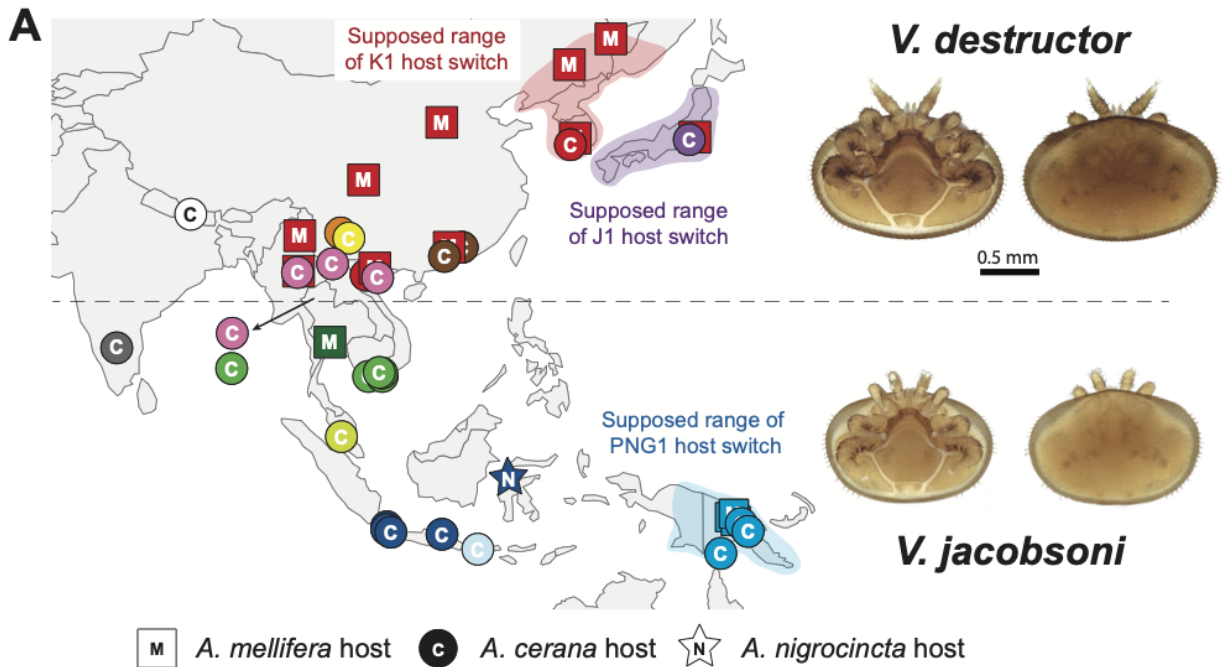


Figure 1: Mitogenomic phylogeographic networks uncover cryptic diversity in host-switched lineages.

Both mite species exhibit geographic structure on their original host, with lineages classified using their *COX1* sequences (A). The color of each circle/square indicates a unique mitogenome sequence and size is proportionate to the sample size (B). Previous studies based on *COX1* found that *V. destructor* host-switched lineages largely belong to the so-called 'Korean K1' haplotype. However, the mitogenome-wide analysis revealed four six mitotypes on the new host, suggesting previously unreported switches by mites from the same geographic region (B). For the most recent jump in 2008, out of the seven lineages identified for *V. jacobsoni* only the known PNG1 succeeded in crossing the species barrier [4] (C). Yet, looking at the whole mitochondrial genome, it becomes evident that the host switch involved a number of independent female founders from closely related populations, or even the same population. Thus, mitogenomic data suggest a relatively diverse population of host-switched mites in both species.

We identified 18 distinct mitogenome sequences (hereafter named mitotypes) for *V. destructor* that clustered into seven lineages (Figure 1A and Table S2). Throughout our Asian sampling sites, all *V. destructor* mitotypes on *A. mellifera* belonged to the Korean K1 lineage, which was also found in one *A. cerana* sample (Figure 1B). We found three additional K1 mitotypes in China that were not previously reported on *A. mellifera* [20] (Table S1). We confirmed that the SNPs differentiating these K1 mitotypes were not artifacts of base calling by manually checking the read alignments. Interestingly, the two publicly available mitochondrial genomes of the K1 haplotype (France: NC_004454.2 and Japan: AP0195523.1) appear slightly divergent from the dominant K1 mitotype. Though we found the other reported host switched lineage (Japanese J1) in the source population from Japan, we did not detect it on *A. mellifera*. In addition, a re-analysis of 22 *COX1* sequences available on NCBI, combined with our new data also indicates diverse mitochondrial backgrounds of host-switched mites, both in *V. destructor* and *V. jacobsoni* (Figure S1).

Despite similar sample sizes for both species, we found a much higher genomic variability in the mtDNA in *V. jacobsoni* samples compared to *V. destructor*. Our analysis detected 27 unique mitotypes that belonged to seven lineages (Figure 1C and Table S2). Unexpectedly, we found that the misidentified mite on *A. mellifera* in Thailand (THV003_1) clustered with a new lineage, VJ North Thailand 4 (Table 1). This lineage was closely related to the

North Thailand 1-3 COX1 reference sequences (Figure S1) on and Laos 1 mitotypes (Figure 1C) [10,13]. All other *V. jacobsoni* mites collected on *A. mellifera* ($N = 10$), belonged to the previously reported host switched lineage Papua New Guinea 1. The median joining network showed six mitotypes found on *A. mellifera* were distinct from each other (up to 8 SNPs difference). We also found the two PNG 1 mitotypes detected during the host switch event (2008) that were retrieved several generations later in *A. mellifera* colonies (2014) (Table S1).

Host-associated genomic differentiation following host shifts

We identified 2,728,471 biallelic SNPs in whole-genome sequences of *Varroa spp.* mites, after quality and coverage bias filtering. These variants revealed the loss of genetic diversity that accompanied the transition from the original to the novel bee host. For instance, a 2.8-fold reduction in polymorphic SNPs numbers was observed on the novel host *V. destructor* ($N_{VdesAcer} = 1,219,990$ vs $N_{VdesAmel} = 429,420$ SNPs) and 3.6 fold in *V. jacobsoni* ($N_{VdesAcer} = 1,546,366$ vs $N_{VdesAmel} = 630,534$ SNPs). Reduction of genetic diversity on the novel host was also observed in the average genomic nucleotide diversity π in both *V. destructor* ($\pi_{VdesAcer} = 0.0020$ vs $\pi_{VdesAmel} = 0.0005$) and *V. jacobsoni* ($\pi_{VjacAcer} = 0.0014$ vs $\pi_{VjacAmel} = 0.0008$). Genome scans using 20kb sliding windows showed that π values in the new host was overall lower than in the original host (Figure S2). Analysis of the genomic heterozygosity revealed extremely high levels of inbreeding coefficient in mite genomes with the lowest value $F = 0.686$ in *V. destructor* and $F = 0.721$ in *V. jacobsoni* (Table S3). However, when comparing *A. mellifera* and *A. cerana* host populations, we found no significant difference in F estimates for *V. destructor* (two-tailed t-test p -value = 0.263) and *V. jacobsoni* (two-tailed t-test p -value = 0.297). Such levels in both species were previously reported using microsatellites [4] and more likely resulting from incestuous *Varroa spp.* mating.

PCA further illustrated the loss of genetic diversity in host-switched mites coupled with genome-wide differentiation (Figure 2AC). Likewise, F_{ST} and D_{xy} pairwise estimates indicated strong differentiation between nuclear genomes from divergent lineages (Table S4). Within each species, high levels of genetic differentiation all along all seven chromosomes were detected between host-adapted lineages (Figure S2). Interestingly, when comparing the host-switched *V. destructor* specimens to native-host samples from the same mitotype (K1) there was far less genetic differentiation ($F_{ST} = 0.004$) compared with any other native-range population ($F_{ST} > 0.6$) (Table S4). Thus, host-switched populations still carry many of the genes present in the population of origin, a

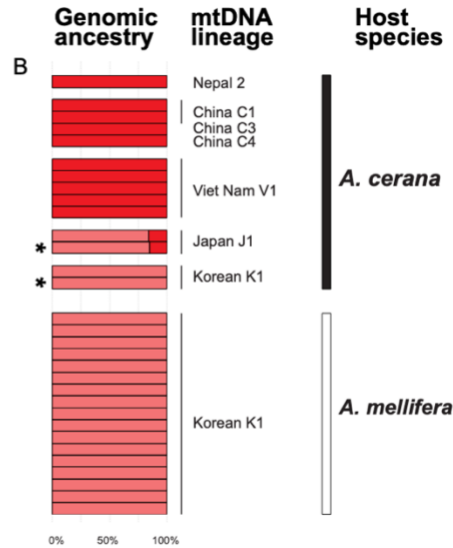
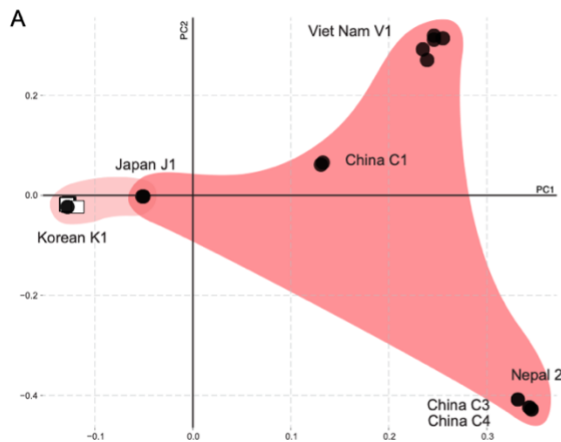
scenario consistent with rapid reproductive isolation post-switch. Similar patterns were obtained genome-wide for the larger sliding windows of 20kb of F_{ST} and D_{xy} estimates show that *A. cerana* and *A. mellifera* K1 mites diverged massively ([Supplemental online markdown](#)).

We found similar patterns of population structure and divergence among hosts and mtDNA lineages in *V. jacobsoni*, while host switch was more recent. In our study, we did not find evidence for quasi-clonality in either *V. jacobsoni* or *V. destructor* mites collected from *A. mellifera* populations.

Ancestry of host switched populations and evidence of gene flow

NGSadmix analysis conducted on 1,276,602 unlinked biallelic SNPs confirmed the population structure observed between *A. mellifera* and *A. cerana* mites and helped to identify the source of host populations. At $K = 2$, population structure quickly emerged between *V. destructor* mites. J1 and K1 *A. cerana* mites were assigned to the same genetic cluster as all K1 *A. mellifera* mites ([Figure 2B](#)). Therefore, genome-wide sequencing confirmed that both K1 and J1 mite lineages contributed to the genetic diversity in host-switched populations. In *V. jacobsoni*, admixture estimates suggested that two large genetic clusters exist: Northern/Asian *V. jacobsoni* lineages differing from the South/Oceanian ([Figure 2D](#)). At $K = 3$, unreported spillover events were seen with the mite (THV003_1) found in *A. mellifera* in Thailand, which was genetically related to mites in Viet Nam (Laos 1). Here, population genomics supports the origin of *V. jacobsoni* host switch in Papua New Guinea to be exclusively from mites infesting *A. cerana* with PNG1 and Java 1 mitochondrial lineages.

V. destructor



V. jacobsoni

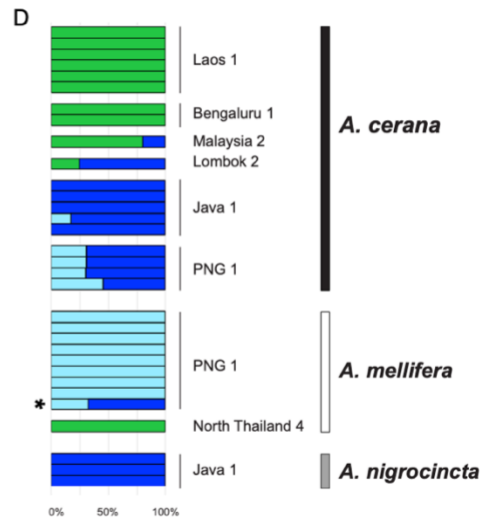
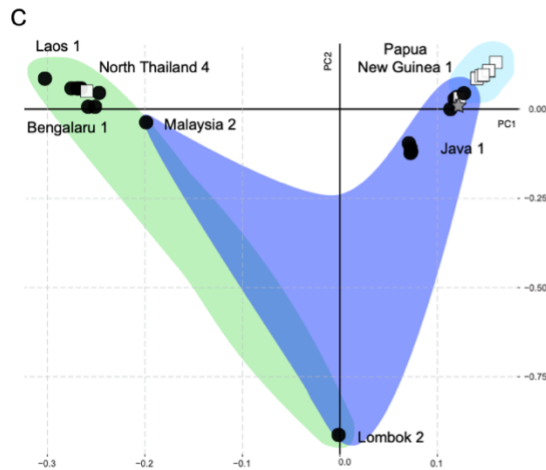


Figure 2: Loss of genetic diversity and rapid genomic differentiation occurred in spite of large founding size and migration between sympatric honey bee hosts. *V. destructor* populations on its original host *A. cerana* (black circle) show geographic structure across their native range (A). Envelopes around the PCA points are color-coded according to nuclear genetic ancestry (B). Host-switched populations (white square) are genetically homogenous throughout Asia and are most similar to the populations with the same mitotype (K1) (B). Despite a more recent host switch (~2008), a similar pattern is observed for *V. jacobsoni* in South-Asia and Oceania, where the PNG1 lineage underwent a host switch (C). However, other switches appear possible

elsewhere in *V. jacobsoni*'s range, such as the specimen collected from *A. mellifera* in North Thailand (green cluster) (D). In both species, there is evidence of mites drifting between hosts, as indicated by genetic analysis of their gut contents, indicated by asterisks. The overall picture for both species is similar, where reproductive differentiation after host switches is coupled with additional opportunities for gene flow via drifting, and potential for additional switches happening elsewhere in the range.

Recent migration and spillover events were confirmed by analyzing mite's bee reads from inside the mite. We first evaluated this method with mite families (used for *de novo* mutation measurements) that were actively feeding at the time of collection and were able to accurately identify the host (Table S5). For other samples up to 0.18% of all reads from each mite library matched honey bee host mitochondrial reference genomes. Using this method we found that a *V. jacobsoni* mite collected on *A. mellifera* that did not cluster with the other host-switched mites based on nuclear DNA (PGV956_4) actually had predominantly *A. cerana* DNA inside her, suggesting a recent drifting event. While admixture estimates did not spot such outliers in *V. destructor*, the diet molecular analysis detected cases in Viet Nam (VNV475_1) and in Japan (JPV025_1) where mites from *A. mellifera* drifted back to *A. cerana*. Since *A. mellifera* and *A. cerana* honey bees present enough important morphological and behavioral differences to make confusing them impossible by experienced field workers, such parallel observations likely indicate that the mites migrated between hosts shortly before collection.

Low *de novo* mutation rate from pedigree-based measurements

We examined whether the rapid parallel differentiation observed between *Varroa spp.* host populations was driven by high frequency of new mutations by estimating the *de novo* mutation rate, following a previously published approach [21]. As the *V. destructor* 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 9x–47x for the mothers and 9x–54x for the sons. Before filtering, the number of detected mutations was 3,082 but there was extreme variation between samples in the number of potential mutations called. Notably mother-son pairs with less coverage contained several folds more mutation calls than mother-son

pairs with higher coverage. For example, between Mom1 (47x) and Son1 (54x), 2 potential mutations were identified, whereas between Mom19 (9x) and Son19 (13x), 1582 potential mutations were identified. Samples with
 205 higher mutation calls also showed stronger evidence of clustering of mutation calls, which typically indicates false-positive mutation calls. For example, between Mom7 (9x) and Son7 (11x), 1454 potential mutations were identified; however, the null hypothesis that mutations were distributed to contigs in proportion to contig lengths was rejected (via a G-test) and the null hypothesis that mutations were distributed uniformly along the contigs was also rejected (via an Anderson-Darling test). Mom2-Son2 and Mom19-Son19 had similar results. After filtering to
 210 remove low quality mutation calls, two mutation calls remained, but showed low coverage and were discarded as unreliable.

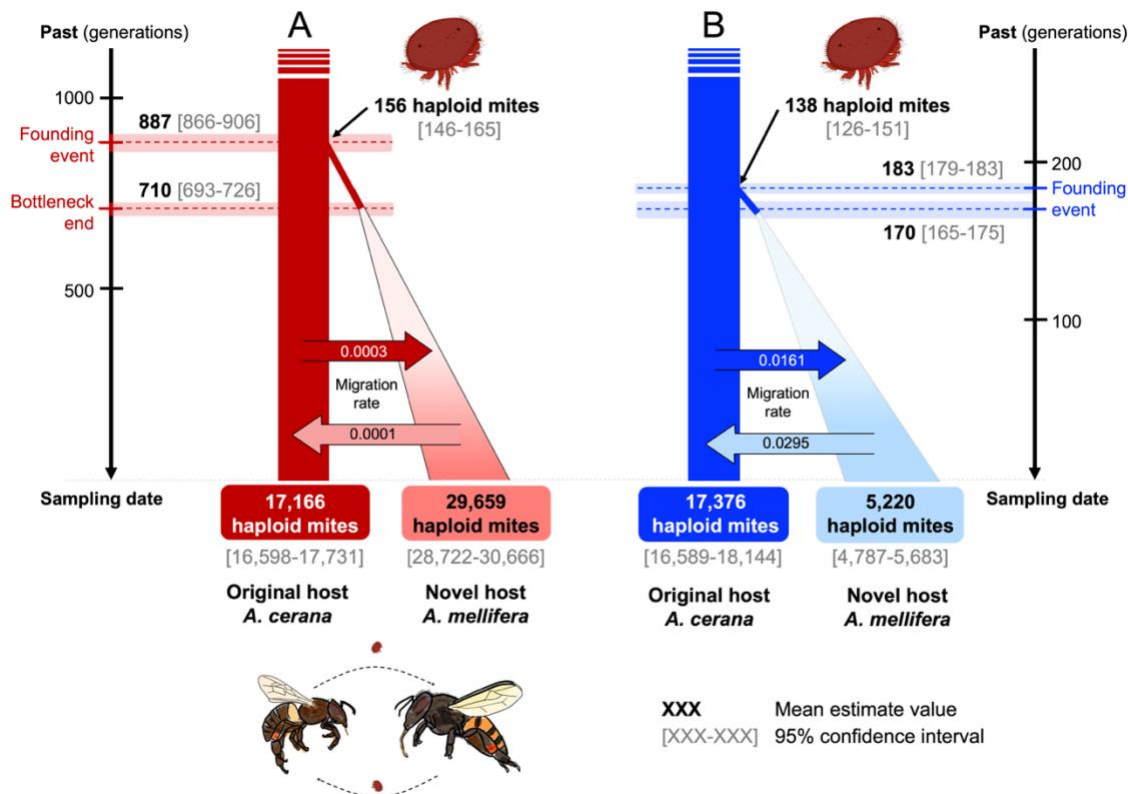
We conducted a set of simulations to deal with coverage bias issues in the data set. Out of 6,000 simulated mutations, 3,464 could be detected by our pipeline, indicating that our recall rate was approximately 58%. The total haploid length of the 7 chromosomes considered in the mutation scan was 362,833,908 base pairs. Given
 215 that we found no mutations among 6 mother-son pairs, we estimate that the mutation rate in *V. destructor* 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.

Estimating of host switch demographic history and founding population size using nuclear DNA

We aimed to further parameterize the demography of host switches detected in mitochondrial data using coalescent-based modeling based on nuclear DNA. Demographic inference using fastsimcoal2 was conducted
 220 using an isolation with migration scenario (Figure S3). For *V. destructor*, 32 haploid mite genomes were sequenced from *A. mellifera* host and 22 haploid mite genomes were sequenced from *A. cerana* host. For *V. jacobsoni*, 18 haploid genomes of *A. mellifera* mites and 6 haploid genomes of *A. cerana* mites from Papua New Guineas were projected (excluding the older sample from 1997). Given that we were only able to estimate the upper limit of *de novo* mutation rate, we explored whether lower mutation rate improve the observed maximum
 225 likelihood distribution with 100 replicate simulations (Figure S4). We examined rates as low as 1.0×10^{-11} , at the lower limit of rates in arthropods [22]. We also tested whether including an inbreeding coefficient would help improving the model likelihood by adding 70% of inbreeding (minimum detected in both species). However, lower

mutation rates did not improve model fit for both *V. destructor* (Kruskal-Wallis chi-squared = 497.61, df = 494, *p*-value = 0.446) and *V. jacobsoni* (Kruskal-Wallis chi-squared = 499, df = 498, *p*-value = 0.479) (Figure S4). The
230 inbreeding coefficient did also not significantly improve the likelihood with equal mutation rate (available in [online supplemental markdown](#)). In addition, 2D-joint SFS visual inspection did not reveal striking differences, and the parameter estimates overall were qualitatively similar over the range of mutation and inbreeding parameters we examined.

Hence, demographic parameters and their bootstraps values were estimated using $\mu = 8 \times 10^{-10}$ per bp per
235 generation and no inbreeding value. Mean values and 95% confidence intervals for each species are graphically summarized in Figure 3. The older host switch with *V. destructor* was estimated to have occurred 887 generations before sampling (~88 years ago assuming 10 generations / year [23]) with a founder effect lasting for 177 generations. Estimates suggested that 156 [CI95 = 146-165] haploid genomes of *V. destructor* contributed to the founding event in Asia. From that point on, the novel mite population grew quickly on *A. mellifera* and diverged
240 from the sympatric *A. cerana* source populations. Our results suggest low but continuous gene flow between hosts. Despite being a more recent event, estimated around 183 generations ago (~18 years ago), demographic parameters for *V. jacobsoni* also supported a large founding size reaching 138 haploid mites. The effective population size of *V. jacobsoni* mites on *A. mellifera* was lower than the source population, in contrast to estimates for *V. destructor*.



Discussion

Virtually every living species has at least one 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 [24,25]. Theoretical modelling links the two scales, suggesting that specialization coupled with trade-offs in performance on alternative hosts should lead to speciation [26]. 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 [27]. For this reason, pandemic spread of *Varroa spp.* among honey bees has long been puzzling, given their previously assumed “quasi-clonal” structure in the invasive portion of their range [9]. Using high-resolution mitochondrial and nuclear 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 [28,29]. 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 [14]. 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 spp.*, 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 ~50 years earlier than *V. jacobsoni*. *Varroa spp.* coevolve quickly with their new hosts [30,31] 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
285 selection acts on many genomic regions [32,33]. By promoting genomic heterogeneity and introducing beneficial
mutations [34], gene flow can ameliorate lost diversity due to selection or inbreeding [35,36], both of which would
impact host-switched *Varroa spp.* 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 ($< 8 \times$
 10^{-10} per bp per generation), compared to other arthropods [22,37]. Consequently, immigration may be the
290 primary means of introducing new alleles into populations on the novel host during early stages of the host switch.
Recent work has shown experimentally that gene flow between can occur between *V. destructor* populations on
the two hosts in the native range [20]. This observation, together with migration between host species detected in
this study provides a mechanism by which gene flow occurs – through occasional co-infestations by females from
different populations. In this case, even if a female is not fully adapted to a host, she may still be able to
295 propagate her genes by laying a male even without mating [38], who can then mate with the other female's
daughters [39,40].

While host switch events by specialized parasites were once assumed to be rare, this assumption is increasingly
challenged [41,42]. Ecological fitting theory suggests that shifts could readily occur in species with a pre-existing
ability to use novel hosts [43]. Host switches in *Varroa spp.* appear to fit this model. While the western honey bee
300 has been introduced throughout the *Varroa spp.* native range, mitochondrial data, which are geographically
informative (Figure 1), suggest that switches occurred only in Korea, Japan, Philippines (*V. destructor*), Papua
New Guinea and possibly Thailand (*V. jacobsoni*) [4,11–13]. *A. cerana* subspecies are strongly differentiated
geographically [44–49], and *Varroa spp.* mitotypes mirror host biogeography and subspecies distribution
[10,12,13,50]. This suggests that *Varroa spp.* populations may vary in traits, such as host specificity, as a result
305 of their previous coevolutionary interaction with local *A. cerana* subspecies. This may have allowed some
populations to switch, while others were unable to do so, in spite of available novel hosts [51]. 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 and temporal sampling, given the size of the
native range. While small sample size is enough to retrace ancestral events, large sample sizes can increase the

310 confidence in model selection and parameter estimates for recent demographic events [52]. Yet, past scenario
and demographic estimates were consistent with mite biology and reflected direct observations made during their
invasion of *A. mellifera*. Furthermore, while our data set captured most of the described *Varroa spp.* mitochondrial
lineages and even discovered new ones [3,9,13,53], some reported mitotypes were missing. In the future,
incorporating larger sample size from modern source populations could help better estimate each lineage genetic
315 contribution to host switch success. In addition it would allow to estimate how host switch may have affected
native diversity as for the case of J1 lineage displacement [54,55]. Despite these limitations, our analysis correctly
reconstructed known aspects of *Varroa spp.* 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 10 generations per
320 year [23]. Finally, the use of historical samples, which were collected over the span of approximately a decade,
may also affect parameter estimates in some way. As a result, we caution against over-relying on the apparent
precision of the numerical estimates (Table 2). Nonetheless, data from both mitogenome sequencing and
coalescent analysis all point to a relatively large founding population size of the *Varroa spp.* mites, the existence
of migration and striking parallels in the demographics of host switches between the two species (Figures 1 and
325 3).

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 [12] and the presence of new haplogroups in Eastern Europe [56],
indicating that populations of *Varroa spp.* are continuously testing *A. mellifera* as a new host, and may have
330 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 [57]. Since *V. jacobsoni* presents striking parallels to *V. destructor*, this
knowledge could be applied to forecast and prevent its spread.

Table 2: Parameter estimates for the most likely demographic scenarios with isolation with migration and population expansion for each species jump. As stated in fastsimcoal2 manual, these upper-bound priors indicated with an asterisk were not bounded by the analysis.

	Priors		<i>V. destructor</i>		<i>V. jacobsoni</i>	
	Distribution	Range	Mean value	CI 95%	Mean value	CI 95%
N_{VAC1} (haploid genomes) At present, mite effective population size on original host <i>A. cerana</i>	uniform	[2; 20,000*]	17,166	[16,598; 17,731]	17,376	[16589; 18144]
N_{VAM0} (haploid genomes) At present, mite effective population size on novel host <i>A. mellifera</i>	uniform	[2; 20,000*]	29,659	[28,722; 30,666]	5,220	[4,787; 5,683]
N_{BOTAM} (haploid genomes) Past mite effective founding population size during host switch	uniform	[2; 1,000*]	156	[146; 165]	138	[126; 151]
T_{JUMP} (generation) Time of new founding population split by host switching	uniform and bounded	[1; 1,000]	887	[866; 906]	183	[179; 187]
T_{BOTEND} (generation) End of the founder effect period	uniform and bounded	[1; 1,000]	710	[693; 726]	170	[165; 175]
G_{AM} (haploid genomes per generation) Growth rate of the expanding newly founded mite population (from present to past)	uniform	[-1; 0*]	-0.0014	[-0.0015; -0.0013]	-0.4636	[-0.4895; -0.4399]
M_{ACToAM} (haploid genomes per generation) Migration rate of parasite spillover (original toward new host)	log uniform	[1x10 ⁻⁶ ; 100*]	0.0003	[0.0003; 0.0003]	0.0161	[0.0154; 0.0168]
M_{AMtoAC} (haploid genomes per generation) Migration rate of parasite spillback (new toward original host)	log uniform	[1x10 ⁻⁶ ; 100*]	0.0001	[0.0001; 0.0001]	0.0295	[0.0283; 0.0309]
Mutation rate (fixed-parameter)	-	8.0 x 10 ⁻¹⁰				
Inbreeding (fixed-parameter)	-	none				

340 **Materials and Methods**

Mite sampling on original and novel honey bee hosts

We sampled and sequenced mites for two purposes: 63 *Varroa spp.* females were collected for population genomics (described in this sub-section) and 12 specimens were used for *de novo* mutation rate estimation (described in a later sub-section). We sequenced adult females throughout the *Varroa spp.* native ranges where
345 original and novel hosts occur in sympatry (for population genomics: *V. destructor* $N = 31$ and *V. jacobsoni* $N = 32$; Table 1). For *V. destructor*, specimens were collected between 1996 and 2003 as part of the taxonomic revision and diversity survey carried by Anderson and Trueman [3] and Navajas *et al.* [11] and included 14 samples from *A. cerana* and 17 from *A. mellifera* collected from Japan, South Korea, China, Thailand, Viet Nam, Myanmar and Nepal (see interactive maps [https://maevatecher.github.io/varroa-host-jump/#sampling-](https://maevatecher.github.io/varroa-host-jump/#sampling-distribution)
350 [distribution](https://maevatecher.github.io/varroa-host-jump/#sampling-distribution)). For *V. jacobsoni*, female specimens were obtained from Anderson and Trueman [3] (1994-2004) and Roberts *et al.* [4] (2008-2015) surveys of host shifts reports in Papua New Guinea. A total of 19 adult female *V. jacobsoni* mites were collected from *A. cerana* across different areas in India, Thailand, Viet Nam, Malaysia, Indonesia and Papua New Guinea. In contrast to *V. destructor*, *V. jacobsoni* infestation on *A. mellifera* is still restricted to Papua New Guinea from which 10 samples were collected. Additionally, we included three *V.*
355 *jacobsoni* mites collected from an alternative original host *A. nigrocincta* found in Indonesia. *A. nigrocincta* is not believed to be the source of the host-switched mites, so these samples were included as an outgroup.

Sampling complete details regarding host, location and year are available in Table S1. Exact geographical coordinates were not always available for samples collected before 2008, and were approximated by the locality provided, or published survey maps [3,11,53,58]. All individuals were mature sclerotized females collected from
360 single colonies and were preserved in individual Eppendorf tubes. Each collection tube was stored with 70% ethanol and kept at -20°C at the CSIRO in Canberra, Australia.

DNA extraction and whole genome resequencing

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

365 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, 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). Libraries were run on HiSeq 4000 and NovaSeq6000 in 150 bp x 2 paired-end mode (Illumina ®) at the OIST Sequencing Center. Biosamples and DRA accession as well as sequencing coverage for each sample are provided in Table S1.

Host DNA contents in mite whole-body metagenomics

We cross-checked honey bee host identities reported during sampling with host read identity to detect potentially migrating mites. Mites feed on honey bees during the phoretic phase [59] and maintain a consistent feeding regimen by consuming ~1 µL of host fluid (digested fat-body and haemolymph) per day [60]. Therefore, we assumed that host DNA would be retrieved from crushed whole-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* [61] and [NC_014295] for *A. cerana* [62]. The number of reads mapped to either one of these honey bee host reference genomes was counted and compared to sampled host identities.

385 Data filtering, mapping and genotype calling

Commands used for each analysis step are available on our Snakemake script [63] available on <https://github.com/MaevaTecher/varroa-host-jump>. Briefly, we assessed demultiplexed fastq read quality using FastQC [64]. We then mapped reads to the *V. destructor* reference genome on NCBI [GCF_002443255.1] [65] separately from the complete mitogenome [NC_004454.2] [66] using the soft-clipping and very sensitive mode of

390 NextGenMap v0.5.0 [67] (following a comparison with Bowtie2 v2.6 [68]). Reads were sorted and duplicates were removed using SAMtools [69], and subsampled to a maximum coverage of 200 using VariantBam [70] to speed up processing. Mapping rates and reads depths were computed from the generated BAM files.

We generated three data sets for the mapped reads depending on analysis requirements, following Yamasaki *et al.* [71]. First, we obtained a “*SNP-only dataset*” containing only variant sites using FreeBayes v1.1.0 [72] with the following parameters: minimum mapping = 10, minimum base quality = 5, use of the four best SNP alleles, and no populations prior. In order to correct for coverage bias and sequencing artefacts in problematic regions, we estimated twice the mean reads depth along the seven main chromosomes. Subsequently, variants were filtered using VCFtools v0.1.12b [73]. The “*SNP-only dataset*” resulted in 2,728,471 SNPs. Second, we computed an “*all-sites dataset*” by calling both variants and invariants sites using BCFtools v1.9 mpileup [74]. We removed 400 indels and sites at over twice the computed mean depth, as well as sites with any missing data. We excluded sites that were not placed on the seven chromosomes. After filtering, the “*all-sites dataset*” resulted in 2,130,335 SNPs and 120,279,163 monomorphic sites. Third, we obtained a “*mtDNA SNP-only dataset*” by calling polymorphic sites only on the mitochondrial genome NC_004454.2. We used FreeBayes v1.1.0 with strict quality parameters. The dataset resulted in 2,091 SNPs. These were validated in the *COX1* region by Sanger 405 sequencing (see next section).

Comparative mitogenomes analysis and lineages identification

We used the “*mtDNA SNP-only dataset*” for mitochondrial variation genomic analysis by converting and generating individual nucleotide sequences using the option *vcf2fasta* in vcflib scripts (github.com/vcflib). Nucleotide sequences were screened and aligned (16,476 bp) using Geneious Prime® 2019.2.3 together with the 410 reference mitochondrial sequences NC_004454.2 (*V. destructor* on *A. mellifera*, France [66]) and also, to avoid problems with reference bias, from AP019523.1 (*V. destructor* on *A. mellifera*, Japan [75]). Mitotype diversity for each *Varroa* species and distribution in hosts was assessed using the median joining network method from SplitsTree5 [76]. Divergence levels between *Varroa* species was estimated by comparing the whole-mitochondrial consensus sequences with Geneious Prime®.

415 Standard methods exist to determine *Varroa* species and lineage (also named haplogroup) identity by using the following mitochondrial markers: 1) a 458-bp fragment from the *COX1* gene [3,77] and 2) a concatenated 2,696-bp fragment including *COX1*, *ATP6*, *COX3*, and *CYTB* [11]. Thus, we determined the lineage (e.g. Korean K1, Japan J1, *etc.*) of each mitotype by extracting the *COX1* barcoding region of interest and aligning it together with 60 unique *COX1* reference sequences from NCBI using ClustalW and manual check. A neighbor-joining tree was
420 computed on the fasta alignment using IQ-tree [78] and exported with iTOL [79]. Following *Varroa* revised nomenclature [5], a sequence was considered from a known lineage only if 100% identical to reference *COX1* haplotype. The same process was used for sub-lineages (e.g. Korean K1-1/K1-2, Korean K1-3,...) identification by with the *COX1-ATP6-COX3-CYTB* concatenated region sequences which were aligned with 22 available reference concatenated sequences for *V. destructor* only [11,56,80].

425 To ensure that filtering of the “*mtDNA SNP-only dataset*” did not remove existing and known variants, we also sequenced the *COX1* gene using Sanger sequencing using primers 10kbCOIF1 and 6,5KbCOIR [11]. 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 10kbCOIF1 and 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
430 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
435 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 differences.

Genome-wide variation and analysis of divergence

We examined levels of genetic diversity and differentiation between host-adapted population of each *Varroa*
440 species. To investigate differentiation among populations, we first measured the absolute divergence D_{xy} which

requires both variants and invariants sites [81]. We also investigated levels of standing genetic variation (π). Genome scans were computed using a sliding window method on the “*all-sites dataset*” (50kb window, sliding every 20kb and containing a minimum of 100 SNPs). This method was implemented using the *parseVCF.py* and *popgenWindow.py* python scripts (github.com/simonhmartin/genomics_general) [82].

445 Using the biallelic “*SNP-only dataset*”, we also estimated the genome-wide differentiation by calculating Weir and Cockerham’s F_{ST} per site using VCFtools. Subsequently, we assessed the genetic structure within species using a principal component analysis (PCA) performed with the R package vcfR [83]. To further investigate the population structure and related ancestry, we conducted an incremental K step analysis using NGSadmix [84]. We reduce the effect of linkage disequilibrium between SNPs by conducting a pruning using PLINK v1.90b3 [85] on the
450 “*SNP-only dataset*” with the following parameters: --indep-pairwise 20 10 0.5. After removing linked SNPs, the pruned dataset containing 1,276,602 SNPs was converted into BEAGLE format and directly used in NGSadmix. A total of 10 replicates per K, from 2 to 20, were run and the model with the highest likelihood was selected for plotting each K. The number of genetic clusters among and within species was determined following guidelines in [86,87] and making biological sense.

455 **Mutation rate estimation from mother-son pairs**

Coalescent analysis requires an estimation of mutation rate, yet none is available for closely related species. To do this we collected and sequenced six *V. destructor* diploid mothers and their haploid sons for this purpose. These samples were collected from individual 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
460 mite, 2) a *V. destructor* family at this stage should 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. We collected three families 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.

465 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.

470 Mutation rate was estimated from GATK variant calls [88] using DeNovoGear's dng-call algorithm [21] which can model *Varroa*'s haplodiploidy sex-determination system. Single-nucleotide mutations were called on the 7 largest contigs in the Vdes_3.0 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) 475 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.

480 **Demographic analyses of host switch using SFS**

We inferred the demographic history of both *V. destructor* and *V. jacobsoni* mites with the coalescent simulator fastsimcoal v2.6 using the site frequency spectrum (SFS) [89]. Demographic inferences were computed independently for each species and samples were subsetted following genome-wide analysis results while considering geographical sympatry and continuity. For *V. destructor* ($N = 27$), we excluded Nepal and the 485 Japanese mites (continental island) whereas for *V. jacobsoni* ($N = 12$), we included only mites from Papua New Guinea (host switch region). To reduce the effect of selection which can bias the SFS [71], we kept SNPs and invariants from the "all-sites dataset" that were at least 50 kb away from any annotated genes regions (VCFtools exclude-bed option). The 2D-joint folded-SFS was computed for each species/host population on the filtered 12,594,802 sites including 224,568 SNPs, using the *vcf2sfs* R scripts (github.com/shenglin-liu/vcf2sfs) [90].

490 We considered a scenario consistent with the known history of *Varroa spp.* jumps ([Figure S3](#)). We incorporated the following demographic parameters ([Table 2](#)): the estimated effective population size for *Varroa* mites parasitizing *A. cerana* N_{VAC1} (in haploid genomes) to be stable before, during, and after the host switch. On the other hand, the mite population size on the new host *A. mellifera* expanded from N_{BOTAM} or not with a growth rate G_{AM} after the host switch to reach modern population size N_{VAM0} (in haploid genomes). The host switch founder event occurred at a time T_{JUMP} (in generations) and ended at a time T_{BOTEND} (in generations). Finally, in a case of bidirectional migrations due to sympatry we estimated M_{AMtoAC}/M_{ACtoAM} , to be the proportion of haploid genomes migrating from one population to another.

We ran 100 replicates using the observed SFS as follow: a minimum of 20 loops (`--minnumloops 20`) and a maximum of 150 loops (e.g., ECM cycles, `--numloops 100`) were performed to estimate the parameters, with one million coalescent simulations per loop (`--numsims 1,000,000`), and a 0.001 minimum relative difference in parameter values estimated by the maximum composite likelihood (`--maxlhood 0.001`). The replicate with the highest likelihood set of estimated parameters was retained for model comparison.

To ensure that the upper limit genome-wide mutation rate $\mu = 8.0 \times 10^{-10}$ previously used was adequate to estimate parameters from the observed SFS, we tested different μ levels. For this we ran scenario 4 in the same conditions as for scenario choice but μ ranging from 8.0×10^{-10} to 1.0×10^{-11} . Additionally, fastsimcoal2.6 offers the possibility to input inbreeding coefficient. As *Varroa spp.* populations showed high inbreeding, we also ran scenario 4 with ($F = 0.7$, preliminary tests also ran with 0.8 and 0.9) and without inbreeding coefficient. For all these conditions combined, we ran 100 replicates for each of the 10 combinations per species.

Finally, the overall best estimate parameters set was used to generate 100 pseudo-observed SFS (for a similar number of polymorphic SNPs) for parametric bootstraps. We repeated the parameters estimation for each pseudo-observed SFS and kept the best run. The top 100 runs estimated parameters values were used to calculate 95% confidence intervals. The mean and 95 percentile confidence intervals were computed using the 'boot' R package [91]. Finally, the fit of the best expected SFS was visualized against the observed using SFStools R scripts (github.com/marqueda/SFS-scripts).

515 **Data Availability**

All sequencing data generated for this study have been deposited in the DNA Databank of Japan and transferred to NCBI under the bioproject PRJDB9195 (<https://www.ncbi.nlm.nih.gov/bioproject/650181>) with the accession series DRR209082-DRR209125 and DRR212369-DRR212380. All custom scripts and pipeline developed using Snakemake, Rmarkdown have been made available on GitHub: <https://github.com/MaevaTecher/varroa-host-jump>. Demographic inferences input files and FASTA sequences alignment (mtDNA) are made available in the same repository. Variant calling files, genome indexing files and input lists are readily available from the following DRYAD repository: <https://doi.org/10.5061/dryad.mgqnk98x1>.

Author Contributions

525 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. JMCR collected and provided samples from the mite CSIRO collection and participated in the data interpretation. RAC analyzed data, wrote the manuscript, provided reproducible online resources and data for the estimation of mutation rate.

530 **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). JMCR was supported by the Australian
535 Centre for International Agricultural Research (ACIAR) and the Australian Department of Agriculture, Water and Environment (DAWE). RAC was supported by the 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.

540

Competing interests

The authors declare that they have no competing interests.

Ethics approval and consent to participate

545 Not applicable.

Consent for publication

Not applicable.

References

- 550 1. Hempel PS. Evolutionary parasitology: the integrated study of infections, immunology, ecology, and genetics. Oxford University Press,; 2011.
2. Hatcher MJ, Dick JTA, Dunn AM. Disease emergence and invasions. *Funct Ecol.* 2012;26:1275–87.
3. Anderson DL, Trueman JW. *Varroa jacobsoni* (Acari: Varroidae) is more than one species. *Exp Appl Acarol.* 2000;24:165–89.
- 555 4. Roberts JMK, Anderson DL, Tay WT. Multiple host shifts by the emerging honeybee parasite, *Varroa jacobsoni*. *Mol Ecol.* Wiley Online Library; 2015;24:2379–91.
5. Traynor KS, Mondet F, de Miranda JR, Techer M, Kowallik V, Oddie MAY, et al. *Varroa destructor*. A Complex Parasite, Crippling Honey Bees Worldwide. *Trends Parasitol.* 2020;36:592–606.
- 560 6. Noël A, Le Conte Y, Mondet F. *Varroa destructor*: how does it harm *Apis mellifera* honey bees and what can be done about it? *Emerg Top Life Sci.* 2020;4:45–57.
7. Crane E. Beekeeping round the World. *Bee World.* Taylor & Francis; 1968;49:113–4.
8. Chantawannakul P, de Guzman LI, Li J, Williams GR. Parasites, pathogens, and pests of honeybees in Asia. *Apidologie* . 2016;47:301–24.
- 565 9. Solignac M, Cornuet J-M, Vautrin D, Le Conte Y, Anderson D, Evans J, 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.* 2005;272:411–9.
10. Warrit N, Smith DR, Lekprayoon C. Genetic subpopulations of *Varroa* mites and their *Apis cerana* hosts in Thailand. *Apidologie.* EDP Sciences; 2006;37:19–30.

- 570 11. Navajas M, Anderson DL, de Guzman LI, Huang ZY, Clement J, Zhou T, et al. New Asian types of *Varroa destructor*: a potential new threat for world apiculture. *Apidologie*. EDP Sciences; 2010;41:181–93.
12. Beaufort AL, Truong TA, Fajardo AC, Dinh TQ, Cervancia C, Moritz RFA. Host Specificity in the Honeybee Parasitic Mite, *Varroa spp.* in *Apis mellifera* and *Apis cerana*. *PLoS One*. 2015;10:e0135103.
13. Dietemann V, Beaufort A, Page P, Yañez O, Buawangpong N, Chantawannakul P, et al. Population genetics of ectoparasitic mites *Varroa spp.* in Eastern and Western honey bees. *Parasitology*. 2019;146:1429–39.
- 575 14. Fry WE. *Phytophthora infestans*: New Tools (and Old Ones) Lead to New Understanding and Precision Management. *Annu Rev Phytopathol*. Annual Reviews; 2016;54:529–47.
15. McMullan M, Rafiqi M, Kaithakottil G, Clavijo BJ, Bilham L, Orton E, et al. The ash dieback invasion of Europe was founded by two genetically divergent individuals. *Nat Ecol Evol*. 2018;2:1000–8.
- 580 16. Zayed A, Constantin SA, Packer L. Successful biological invasion despite a severe genetic load. *PLoS One*. 2007;2:e868.
17. Roman J, Darling JA. Paradox lost: genetic diversity and the success of aquatic invasions. *Trends Ecol Evol*. Elsevier; 2007;22:454–64.
18. Mikheyev AS. History, genetics and pathology of a leaf-cutting ant introduction: a case study of the Guadeloupe invasion. *Biol Invasions*. Springer Netherlands; 2008;10:467–73.
- 585 19. Schrieber K, Lachmuth S. The Genetic Paradox of Invasions revisited: the potential role of inbreeding × environment interactions in invasion success. *Biol Rev Camb Philos Soc*. 2017;92:939–52.
20. Lin Z, Wang S, Neumann P, Chen G, Page P, Li L, et al. Population genetics and host specificity of *Varroa destructor* mites infesting eastern and western honeybees. *J Pest Sci [Internet]*. 2021; Available from: <https://doi.org/10.1007/s10340-020-01322-7>
- 590 21. Ramu A, Noordam MJ, Schwartz RS, Wuster A, Hurles ME, Cartwright RA, et al. DeNovoGear: de novo indel and point mutation discovery and phasing. *Nat Methods*. 2013;10:985–7.
22. Fazalova V, Nevado B. Low Spontaneous Mutation Rate and Pleistocene Radiation of Pea Aphids. *Mol Biol Evol*. 2020;37:2045–51.
- 595 23. Cornuet JM, Beaumont MA, Estoup A, Solignac M. Inference on microsatellite mutation processes in the invasive mite, *Varroa destructor*, using reversible jump Markov chain Monte Carlo. *Theor Popul Biol*. 2006;69:129–44.
24. Poulin R, Krasnov BR, Mouillot D, Thieltges DW. The comparative ecology and biogeography of parasites. *Philos Trans R Soc Lond B Biol Sci*. 2011;366:2379–90.
- 600 25. Penczykowski RM, Laine A-L, Koskella B. Understanding the ecology and evolution of host-parasite interactions across scales. *Evol Appl*. 2016;9:37–52.

26. Sexton JP, Montiel J, Shay JE, Stephens MR, Slatyer RA. Evolution of Ecological Niche Breadth. *Annu Rev Ecol Evol Syst.* Annual Reviews; 2017;48:183–206.
27. McDonald BA, Linde C. Pathogen population genetics, evolutionary potential, and durable resistance. *Annu Rev Phytopathol.* 2002;40:349–79.
- 605 28. Auger-Rozenberg M-A, Boivin T, Magnoux E, Courtin C, Roques A, Kerdelhué C. Inferences on population history of a seed chalcid wasp: invasion success despite a severe founder effect from an unexpected source population. *Mol Ecol.* 2012;21:6086–103.
29. Demastes JW, Hafner DJ, Hafner MS, Light JE, Spradling TA. Loss of genetic diversity, recovery and allele surfing in a colonizing parasite, *Geomydoecus aurei*. *Mol Ecol.* 2019;28:703–20.
- 610 30. Beaurepaire AL, Moro A, Mondet F, Le Conte Y, Neumann P, Locke B. Population genetics of ectoparasitic mites suggest arms race with honeybee hosts. *Sci Rep.* 2019;9:11355.
31. Moro A, Blacquière T, Panziera D, Dietemann V, Neumann P. Host-Parasite Co-Evolution in Real-Time: Changes in Honey Bee Resistance Mechanisms and Mite Reproductive Strategies. *Insects.* Multidisciplinary Digital Publishing Institute; 2021;12:120.
- 615 32. Pinho C, Hey J. Divergence with Gene Flow: Models and Data. *Annu Rev Ecol Evol Syst.* Annual Reviews; 2010;41:215–30.
33. Feder JL, Egan SP, Nosil P. The genomics of speciation-with-gene-flow. *Trends Genet.* 2012;28:342–50.
34. Tomasini M, Peischl S. When does gene flow facilitate evolutionary rescue? *Evolution.* 2020;74:1640–53.
35. Lenormand T. Gene flow and the limits to natural selection. *Trends Ecol Evol.* Elsevier; 2002;17:183–9.
- 620 36. Whiteley AR, Fitzpatrick SW, Funk WC, Tallmon DA. Genetic rescue to the rescue. *Trends Ecol Evol.* 2015;30:42–9.
37. Liu H, Jia Y, Sun X, Tian D, Hurst LD, Yang S. 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.* 2017;34:119–30.
- 625 38. Häußermann CK, Giacobino A, Munz R, Ziegelmann B, Palacio MA, Rosenkranz P. Reproductive parameters of female *Varroa destructor* and the impact of mating in worker brood of *Apis mellifera*. *Apidologie.* 2020;51:342–55.
39. Beaurepaire AL, Krieger KJ, Moritz RFA. Seasonal cycle of inbreeding and recombination of the parasitic mite *Varroa destructor* in honeybee colonies and its implications for the selection of acaricide resistance. *Infect Genet Evol.* Elsevier; 2017;50:49–54.
- 630 40. Beaurepaire AL, Ellis JD, Krieger KJ, Moritz RFA. Association of *Varroa destructor* females in multiply infested cells of the honeybee *Apis mellifera*. *Insect Sci.* 2019;26:128–34.

41. Rigaud T, Perrot-Minnot M-J, Brown MJF. Parasite and host assemblages: embracing the reality will improve our knowledge of parasite transmission and virulence. *Proc Biol Sci.* rsos.royalsocietypublishing.org; 2010;277:3693–702.
42. Nylin S, Agosta S, Bensch S, Boeger WA, Braga MP, Brooks DR, et al. Embracing Colonizations: A New Paradigm for Species Association Dynamics. *Trends Ecol Evol.* 2018;33:4–14.
43. Araujo SBL, Braga MP, Brooks DR, Agosta SJ, Hoberg EP, von Hartenthal FW, et al. Understanding Host-Switching by Ecological Fitting. *PLoS One.* journals.plos.org; 2015;10:e0139225.
44. Radloff SE, Hepburn C, Randall Hepburn H, Fuchs S, Hadisoelilo S, Tan K, et al. Population structure and classification of *Apis cerana*. *Apidologie.* EDP Sciences; 2010;41:589–601.
45. Tan K, Qu Y, Wang Z, Liu Z, Engel MS. Haplotype diversity and genetic similarity among populations of the Eastern honey bee from Himalaya-Southwest China and Nepal (Hymenoptera: Apidae). *Apidologie* . Springer Paris; 2016;47:197–205.
46. Chen C, Wang H, Liu Z, Chen X, Tang J, Meng F, et al. Population Genomics Provide Insights into the Evolution and Adaptation of the Eastern Honey Bee (*Apis cerana*). *Mol Biol Evol.* academic.oup.com; 2018;35:2260–71.
47. Yancan L, Tianle C, Yunhan F, Delong L, Guizhi W. Population genomics and morphological features underlying the adaptive evolution of the eastern honey bee (*Apis cerana*). *BMC Genomics.* 2019;20:869.
48. Ilyasov RA, Youn HG, Lee M-L, Kim KW, Proshchalykin MY, Lelej AS, et al. Phylogenetic Relationships of Russian Far-East *Apis cerana* with Other North Asian Populations. *Journal of Apicultural Science.* Berlin: Sciendo; 2019;63:289–314.
49. Ji Y, Li X, Ji T, Tang J, Qiu L, Hu J, et al. Gene reuse facilitates rapid radiation and independent adaptation to diverse habitats in the Asian honeybee. *Sci Adv.* American Association for the Advancement of Science; 2020;6:eabd3590.
50. Rueppell O, Hayes AM, Warrit N, Smith DR. Population structure of *Apis cerana* in Thailand reflects biogeography and current gene flow rather than Varroa mite association. *Insectes Soc.* SP Birkhäuser Verlag Basel; 2011;58:445–52.
51. Li W, Wang C, Huang ZY, Chen Y, Han R. Reproduction of Distinct *Varroa destructor* Genotypes on Honey Bee Worker Brood. *Insects.* Multidisciplinary Digital Publishing Institute; 2019;10:372.
52. Robinson JD, Coffman AJ, Hickerson MJ, Gutenkunst RN. Sampling strategies for frequency spectrum-based population genomic inference. *BMC Evol Biol.* 2014;14:254.
53. Zhou T, Anderson DL, Huang ZY, Huang S, Yao J, Ken T, et al. Identification of Varroa mites (Acari: Varroidae) infesting *Apis cerana* and *Apis mellifera* in China. *Apidologie.* EDP Sciences; 2004;35:645–54.
54. de Guzman LI, Rinderer TE, Anthony Stelzer J. Occurrence of two genotypes of *Varroa jacobsoni* Oud. in North America. *Apidologie.* EDP Sciences; 1999;30:31–6.

55. Ogihara MH, Yoshiyama M, Morimoto N, Kimura K. Dominant honeybee colony infestation by *Varroa destructor* (Acari: Varroidae) K haplotype in Japan. *Appl Entomol Zool.* 2020;55:189–97.
56. Gajic B, Radulovic Z, Stevanovic J, Kulisic Z, Vucicevic M, Simeunovic P, et al. Variability of the honey bee mite *Varroa destructor* in Serbia, based on mtDNA analysis. *Exp Appl Acarol.* 2013;61:97–105.
57. Evans JD, Cook SC. Genetics and physiology of Varroa mites. *Curr Opin Insect Sci.* 2018;26:130–5.
58. Anderson DL, Fuchs S. Two genetically distinct populations of *Varroa jacobsoni* with contrasting reproductive abilities on *Apis mellifera*. *J Apic Res.* Taylor & Francis; 1998;37:69–78.
59. Ramsey SD, Ochoa R, Bauchan G, Gulbranson C, Mowery JD, Cohen A, et al. *Varroa destructor* feeds primarily on honey bee fat body tissue and not hemolymph. *Proc Natl Acad Sci U S A. National Academy of Sciences;* 2019;116:1792–801.
60. Posada-Florez F, Sonenshine DE, Egekwu NI, Rice C, Lupitskyy R, Cook SC. Insights into the metabolism and behaviour of *Varroa destructor* mites from analysis of their waste excretions. *Parasitology.* 2019;146:527–32.
61. Crozier RH, Crozier YC. The mitochondrial genome of the honeybee *Apis mellifera*: complete sequence and genome organization. *Genetics.* Genetics Soc America; 1993;133:97–117.
62. Tan H-W, Liu G-H, Dong X, Lin R-Q, Song H-Q, Huang S-Y, et al. The complete mitochondrial genome of the Asiatic cavity-nesting honeybee *Apis cerana* (Hymenoptera: Apidae). *PLoS One.* 2011;6:e23008.
63. Köster J, Rahmann S. Snakemake--a scalable bioinformatics workflow engine. *Bioinformatics.* 2012;28:2520–2.
64. Andrews S, Others. FastQC: a quality control tool for high throughput sequence data. 2010;
65. Techer MA, Rane RV, Grau ML, Roberts JMK, Sullivan ST, Liachko I, et al. Divergent evolutionary trajectories following speciation in two ectoparasitic honey bee mites. *Commun Biol.* 2019;2:357.
66. Navajas M, Le Conte Y, Solignac M, Cros-Arteil S, Cornuet J-M. The complete sequence of the mitochondrial genome of the honeybee ectoparasite mite *Varroa destructor* (Acari: Mesostigmata). *Mol Biol Evol.* Oxford University Press; 2002;19:2313–7.
67. Sedlazeck FJ, Rescheneder P, von Haeseler A. NextGenMap: fast and accurate read mapping in highly polymorphic genomes. *Bioinformatics. academic.oup.com;* 2013;29:2790–1.
68. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 2012;9:357–9.
69. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 2009;25:2078–9.
70. Wala J, Zhang C-Z, Meyerson M, Beroukhir R. VariantBam: filtering and profiling of next-generational sequencing data using region-specific rules. *Bioinformatics. academic.oup.com;* 2016;32:2029–31.

71. Yamasaki YY, Kakioka R, Takahashi H, Toyoda A, Nagano AJ, Machida Y, et al. Genome-wide patterns of divergence and introgression after secondary contact between *Pungitius* sticklebacks. *Philos Trans R Soc Lond B Biol Sci.* 2020;375:20190548.
72. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing [Internet]. arXiv [q-bio.GN]. 2012. Available from: <http://arxiv.org/abs/1207.3907>
73. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. *Bioinformatics.* 2011;27:2156–8.
74. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics.* 2011;27:2987–93.
75. Harada R, Yoshioka M, Okuyama H, Kato M, Martin SJ, Takahashi J-I. Complete mitochondrial DNA sequence of the parasitic honey bee mite *Varroa destructor* (Mesostigmata: Varroidae). *Mitochondrial DNA B Resour.* Taylor & Francis; 2020;5:635–6.
76. Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol.* 2006;23:254–67.
77. Dietemann V, Nazzi F, Martin SJ, Anderson DL, Locke B, Delaplane KS, et al. Standard methods for varroa research. *J Apic Res.* Taylor & Francis; 2013;52:1–54.
78. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 2015;32:268–74.
79. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* 2019;47:W256–9.
80. Elbeaino T, Daher-Hjaij N, Ismaeil F, Mando J, Khaled BS, Kubaa RA. 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.* 2016;69:11–9.
81. Cruickshank TE, Hahn MW. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Mol Ecol.* 2014;23:3133–57.
82. Martin SH, Davey JW, Jiggins CD. Evaluating the use of ABBA-BABA statistics to locate introgressed loci. *Mol Biol Evol.* 2015;32:244–57.
83. Knaus BJ, Grünwald NJ. vcfr: a package to manipulate and visualize variant call format data in R. *Mol Ecol Resour.* Wiley Online Library; 2017;17:44–53.
84. Skotte L, Korneliussen TS, Albrechtsen A. Estimating individual admixture proportions from next generation sequencing data. *Genetics.* 2013;195:693–702.
85. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81:559–75.

86. Meirmans PG. Seven common mistakes in population genetics and how to avoid them. *Mol Ecol*. 2015;24:3223–31.
87. Lawson DJ, van Dorp L, Falush D. A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. *Nat Commun*. 2018;9:3258.
- 735 88. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20:1297–303.
89. Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M. Robust demographic inference from genomic and SNP data. *PLoS Genet*. 2013;9:e1003905.
- 740 90. Liu S, Ferchaud A-L, Grønkjaer P, Nygaard R, Hansen MM. Genomic parallelism and lack thereof in contrasting systems of three-spined sticklebacks. *Mol Ecol*. 2018;27:4725–43.
91. Canty A, Ripley B. Package ‘boot’ [Internet]. Version; 2019. Available from: <ftp://ftp.math.ethz.ch/sfs/R-CRAN/web/packages/boot/boot.pdf>

Varroa mite species

- *V. destructor*
- *V. jacobsoni*
- *V. underwoodi*
- *V. rindereri*
- xxx Detected in this study
- xxx this study

Honey bee host

- *A. cerana*
- *A. mellifera*
- × *A. nigrocincta*
- ★ *A. dorsata*
- *A. koschevnikovi*

Korean-like haplotypes reported on *A. mellifera* (in and out-of-Asia)

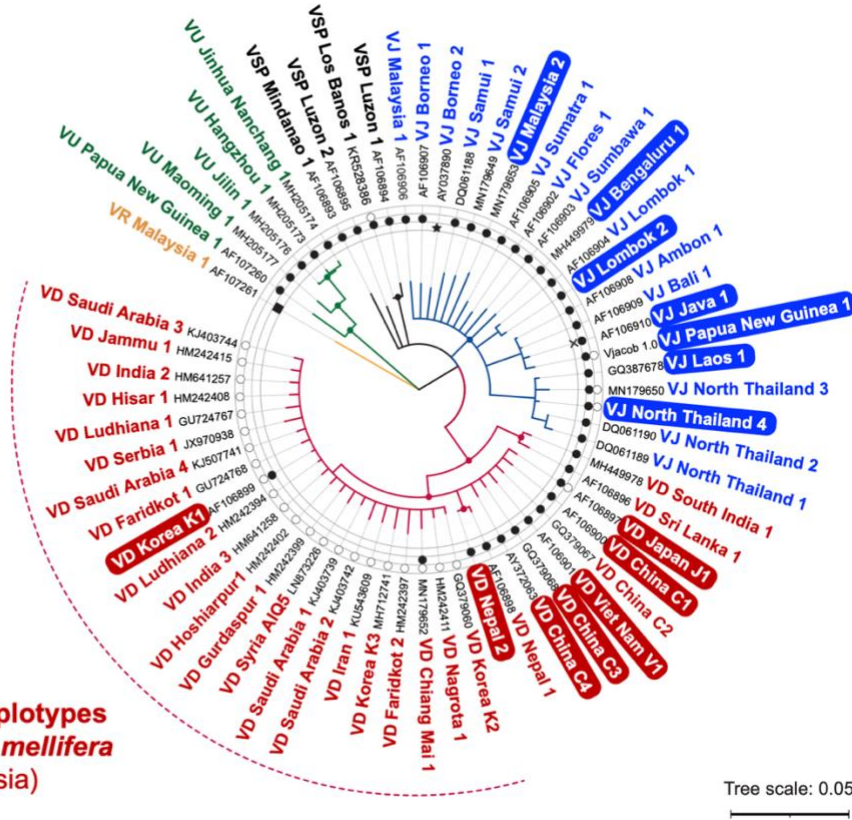


Figure S1: Sequenced *V. destructor* and *V. jacobsoni* mites originated from diverse maternal genetic backgrounds (lineages) shown by mtDNA COX1 barcoding. A neighbor-joining tree was built using the Tamura-Nei genetic distance model with 1,000 bootstraps (> 80% are shown as filled circles in the base of each branch). *V. rindereri* was considered as an outgroup and the host range indicated [3]. *V. destructor* mites collected throughout Asia on their original host belong to seven distinct lineages (see details in Table S2). By contrast, all the host-switched mites across that range belong to the Korean K1 haplotype [AF106899], which is also found worldwide [5]. *V. jacobsoni* mites collected throughout Southeast Asia and Oceania, we identified seven lineages including two newly described: Lombok 2 and North Thailand 4. All mites parasitizing *A. mellifera* in Papua New Guinea were infested only by the *V. jacobsoni* Papua New Guinea 1 haplotype. The novel haplotype, North Thailand 4, was found on one *A. mellifera* colony in the northern limit of *V. jacobsoni* native range.

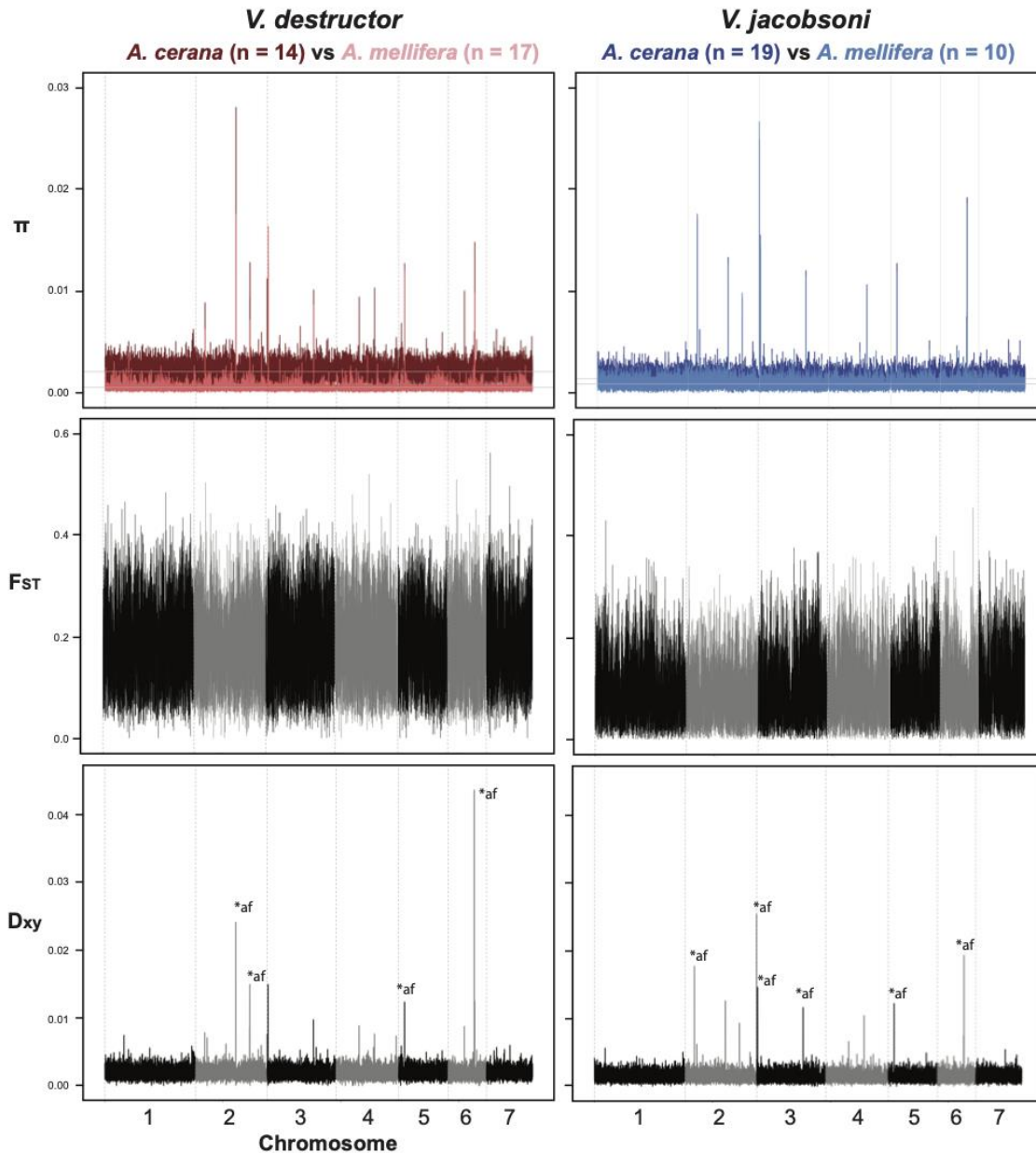


Figure S2: Parallel reduction of genetic diversity and genomic divergence between hosts in *V. destructor* and *V. jacobsoni*. First row: Nucleotide diversity in 20kb window along all seven chromosomes with mites parasitizing the same host pooled as one population for each *Varroa* species. Genetic divergence estimated by pairwise F_{ST} (second row) and absolute divergence using D_{xy} estimate (third row) in 20kb window, show a strong host differentiation all along the genome with few outliers. All sites with D_{xy} value > 0.01 were investigated manually by examining individual genotypes, which showed an excess in heterozygosity levels. These regions were also poorly conserved within species and had BLAST matches elsewhere in the genome, suggesting the presence of sequencing artefacts rather than biological signal (labelled *af on the plots).

Scenario isolation with migration

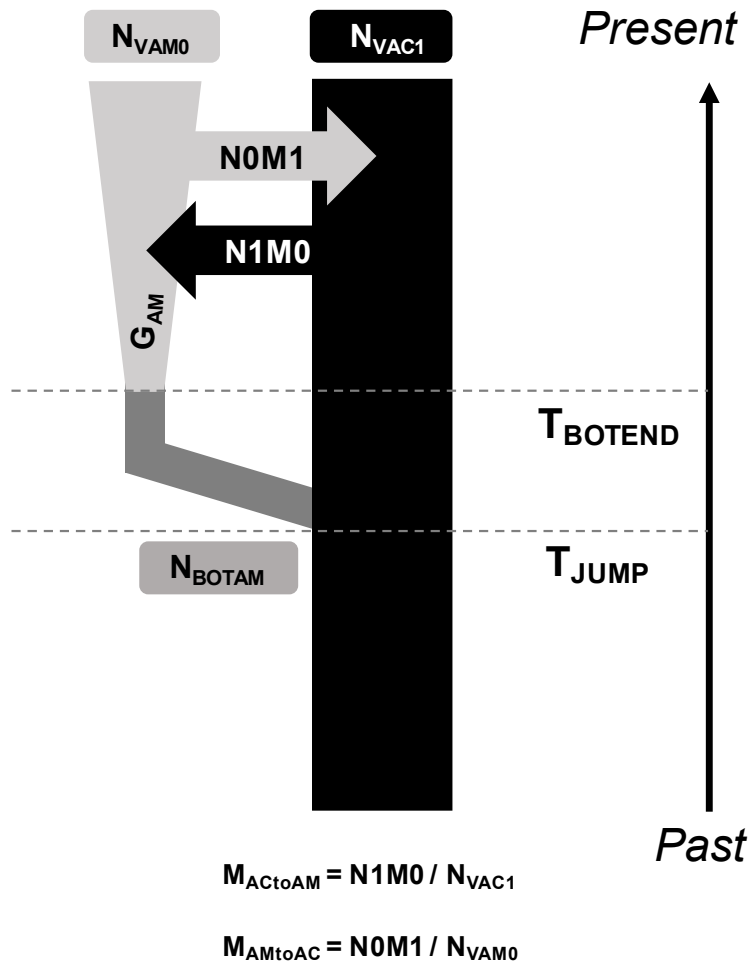
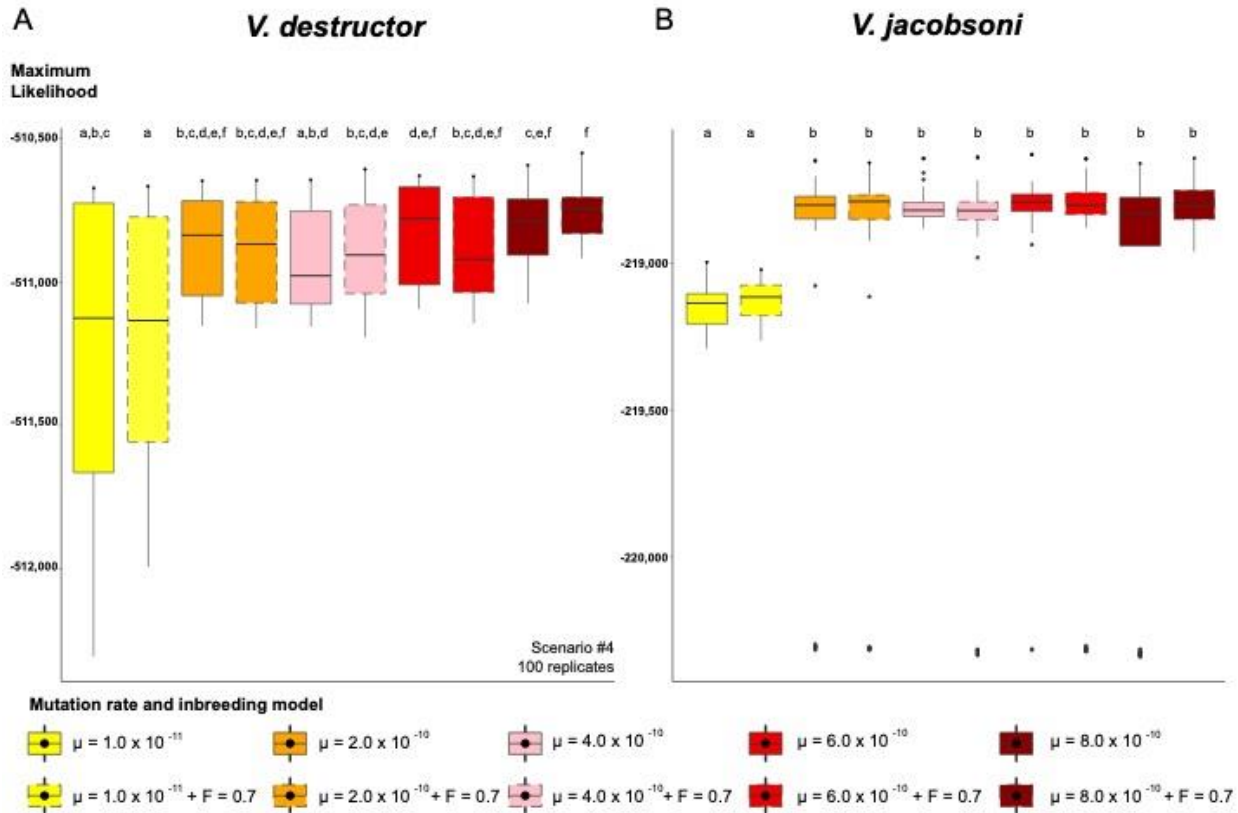


Figure S3 Graphical representation of *Varroa* host switch demographic scenario designed

following population genomics analysis. The model was run independently for each *Varroa* species but contained the same parameters. N_{VAM0} is the present effective population size on the novel host (*A. mellifera*) whereas N_{VAC1} is the present population size on the original host (*A. cerana*). At a time in the past, T_{JUMP} , a founding size of N_{BOTAM} switched durably host from *A. cerana* to *A. mellifera*. The founder effect associated with host switch ended at T_{BOTEND} and population on novel host expanded at a growth rate G_{AM} . As both hosts are in artificial sympatry with evidence of gene flow, migration rates were estimated for spillover (M_{ACtoAM}) and spillback (M_{AMtoAC}). See Table 2 for parameters prior details.



780 **Figure S4 Distribution of the maximum likelihood obtained from the best pseudo-observed SFS**
generated by demographic inferences across different mutation rates and inbreeding levels in
***Varroa* species.** Replicates generated with lower mutation rates than the directly estimated 8.0×10^{-10}
 were not associated with a better fit to the observed SFS from our genomic data for *V. destructor* (A)
 and *V. jacobsoni* (B). No significant differences were detected in the likelihood variance among different
 785 mutation rates. However, pairwise Wilcoxon test significance (letters) after Bonferroni corrections
 showed that the lower mutation rate differentiated from other rates, especially in *V. jacobsoni*.

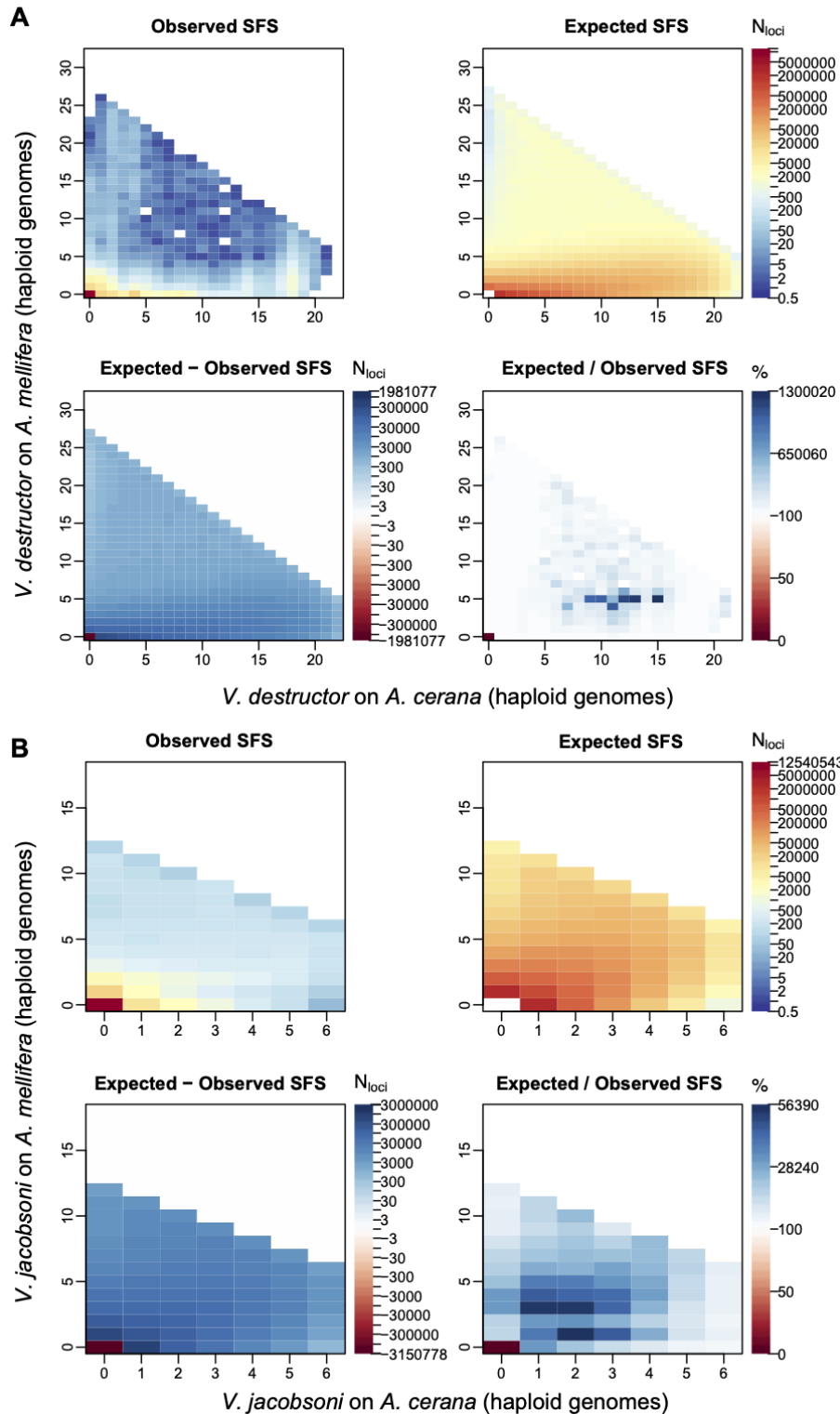


Figure S5: Observed vs expected joint 2D-SFS from the best replicate of the isolation with migration scenario for *V. destructor* and *V. jacobsoni*.

795 **Table S1: Detailed sampling information for each of the *Varroa* mite sequenced for population genomics as well as for estimation of the mutation rate.** Sampling details include *Varroa* species, host identity, country and locality, date and geographical coordinates. Sequencing details include the DRA accession on NCBI, the total number of reads, mapped reads, mapping ratio and mean read depth. Mitochondrial lineages and mitotype identity are also included for each specimen.

800 **Table S2: Repertoire of standard mtDNA markers used to determine mite species and lineages of sequenced specimens.** Names given to new mitotype were chosen to follow the revised nomenclature and compared to *V. destructor* and *V. jacobsoni* reference sequences (source and sequence accession indicated).

Table S3: Observed (O_{HOM}) and expected (E_{HOM}) homozygosity and inbreeding coefficient (F) values computed for each sequenced mite on N_{SITES} for the population genomics analysis.

805 **Table S4: Pairwise Weir and Cockerham F_{ST} values computed between host-adapted *Varroa* species and between host-adapted *Varroa* mtDNA lineages.** The color shades indicate relatively high genetic differentiation (in red), moderate (in yellow) to a low one (in green).

810 **Table S5: Number of reads from whole-body *Varroa* sample mapped against *A. mellifera* and *A. cerana* mitochondrial reference genomes.** Host metagenomics identity was assigned by the higher number of reads and compared to the field collection information. The proof of concept was confirmed with Mom and Son samples collected while feeding on their *A. mellifera* host.