

The ash dieback invasion of Europe was founded by two individuals from a native population with huge adaptive potential

Mark McMullan^{*1}, Maryam Rafiqi², Gemy Kaithakottil¹, Bernardo Clavijo¹, Lorelei Bilham³, Elizabeth Orton³, Lawrence Percival-Alwyn¹, Ben J. Ward^{1,4}, Anne Edwards³, Diane G.O. Saunders³, Gonzalo Garcia¹, Jonathan Wright¹, Walter Verweij¹, Georgios Koutsovoulos⁵, Kentaro Yoshida^{6,7}, Tsuyoshi Hosoya⁸, Louisa Williamson⁹, Philip Jennings⁹, Renaud Ioos,¹⁰ Claude Husson¹¹, Ari M. Hietala¹², Adam Vivian-Smith¹², Halvor Solheim¹², Dan MaClean⁶, Christine Fosker¹, Neil Hall¹, James K.M. Brown³, David Swarbreck¹, Mark Blaxter^{5,13}, Allan Downie³ & Matthew D. Clark^{*1,4}

1. The Earlham Institute Norwich Research Park, Norwich, NR4 7UZ UK, UK
2. Royal Botanic Gardens, Kew, Richmond, Surrey, TW9 3AE, UK
3. John Innes Centre, Norwich Research Park, Norwich, NR4 7UH, UK
4. School of Environmental Sciences, University of East Anglia, Norwich, NR4 7TJ, U.K.
5. Institute of Evolutionary Biology, The University of Edinburgh, Edinburgh, EH9 3FL, UK
6. The Sainsbury Laboratory, Norwich Research Park, Norwich, NR4 7UH, UK.
7. Graduate school of Agricultural Science, Kobe University, Kobe, Hyogo, Japan
8. Department of Botany, National Museum of Nature and Science, Tsukuba, Ibaraki, 305-0005 Japan
9. Fera Science Limited, Sand Hutton, York, YO41 1LZ, UK
10. ANSES Laboratoire de la Santé des Végétaux, 54220 Malzéville, France
11. UMR IAM, INRA, Université de Lorraine, 54000 Nancy, France
12. Norwegian Institute of Bioeconomy Research, Høgskoleveien 8, 1433 Ås, Norway
13. Edinburgh Genomics, School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3JT

*Corresponding Authors:

Mark McMullan (EI) <Mark.McMullan@earlham.ac.uk>

Matthew D. Clark (EI) <Matt.Clark@earlham.ac.uk>

Abstract

Accelerating international trade and climate change make pathogen spread an increasing concern. *Hymenoscyphus fraxineus*, the causal agent of ash dieback is one such pathogen, moving across continents and hosts from Asian to European ash. Most European common ash (*Fraxinus excelsior*) trees are highly susceptible to *H. fraxineus* although a small minority (~5%) evidently have partial resistance to dieback. We have assembled and annotated a draft of the *H. fraxineus* genome which approaches chromosome scale. Pathogen genetic diversity across Europe, and in Japan, reveals a tight bottleneck into Europe, though a signal of adaptive diversity remains in key host interaction genes (effectors). We find that the European population was founded by two divergent haploid individuals. Divergence between these haplotypes represents the ‘shadow’ of a large source population and subsequent introduction would greatly increase adaptive potential and the pathogen’s threat. Thus, EU wide biological security measures remain an important part of the strategy to manage this disease.

Human movement and climate change may increase the incidence of plant pathogen introductions as new environments become favourable¹. Dutch elm disease is one example of an emerging pathogen epidemic which caused the loss of billions of elm trees globally². More recently, ash dieback caused by the ascomycete fungus, *Hymenoscyphus fraxineus*, is causing severe dieback symptoms and death in European common ash *Fraxinus excelsior* and *F. angustifolia* (narrow-leaved ash). *H. fraxineus* jumped host from ash species native to Asia to *F. excelsior*^{3,4}. In its native range, *H. fraxineus* is a leaf pathogen⁵ with little impact on its host, whereas in Europe it is killing European ash at an alarming rate and displacing the non-aggressive indigenous fungus *H. albidus*^{6,7}. The likelihood that a pathogen can infect two plant species increases with their phylogenetic relatedness⁸ and so, host jump epidemics like ash dieback, are an expected risk of global trade of plant material^{9,10}. In line with the propagule pressure theory¹¹, *H. fraxineus* is a typical invasive species that relies on high ascospore production¹².

The disease was first observed in Europe in north-western Poland in 1992 and, moving West, was identified in the UK in 2012^{4,13,14}. The majority of ash trees are susceptible to the pathogen and less than 5% of trees are resistant or tolerant¹⁵⁻¹⁷. Ash dieback disease is characterised by dark brown/orange lesions on leaves followed by wilting, necrotic lesions on shoots then diamond-shaped lesions on the stems and finally preceding dieback of the crown^{13,15,18}. The loss of leaves in the crown of mature trees proceeds over years and leads, in severe cases, to tree death. *H. fraxineus* is a heterothallic fungus¹⁹, shown to reproduce asexually *in vitro*²⁰ but in the wild, probably through annual obligate sexual reproduction on fallen leaf rachises in the leaf litter^{3,21}.

A pathogen's evolutionary potential, is rooted in its genetic diversity (i.e. its effective population size)²². Large populations can adapt more quickly than small ones for two reasons; first, a large population carries more polymorphism and so there is a greater chance that a favourable mutation is present in that population. Second, the impact of random genetic drift is lower for larger populations and therefore, natural selection is better able to drive those favourable mutations to fixation²³. However, pathogen introductions are often associated with genetic bottlenecks which reduce the level of genetic diversity and the efficacy of selection²⁴. This disparity between reduced polymorphism and invasion success is known as the genetic

paradox of biological invasions²⁵. The potato late blight pathogen, *Phytophthora infestans*, is one such example of a successful bottlenecked pathogen introduction²⁶. However devastating a pathogen invasion may be, multiple introductions can increase diversity higher than that of native populations²⁷. Early *P. infestans* invasions were dominated by a single clonal lineage (US-1) but this lineage was superseded by lineages of increasing diversity and increasing severity²⁸. The Dutch elm disease pandemic(s) was also characterised by the initial intense spread of *Ophiostoma ulmi* which later declined and was replaced by *O. novo-ulmi* (also subsp. *americana*), which went on to kill elms across two continents². The level of variation that has been introduced to Europe and the potential for further introduction are of critical concern and a key focus of the present work.

Estimates of *H. fraxineus* microsatellite allelic richness suggest as few as two haploid individuals may have invaded Europe³. Such a bottleneck would represent a genetic paradox and so it is important that we measure nucleotide diversity across the genome as well as meaningful, adaptive diversity in the host interaction genes (effectors). Furthermore, given the impact of the pathogen invasion so far, we set out to understand the characteristics of the source population and the signal of the introduction in order to highlight the consequences of further introduction of polymorphism via identifiable routes of intercontinental disease transmission. Here, we assembled and annotated a high-quality draft genome. We quantified the level of genetic variation in 43 *H. fraxineus* isolates from across Europe as well as 15 isolates from part of its native range (Japan). In order to understand the adaptive potential of the ash dieback pathogen in Europe today as well as its future potential through successive invasion, we determined the effective population size, quantified the bottleneck into Europe and the estimated the size of the source population. In addition, we measured signals of adaptive diversity in genes, including the secreted protein genes encoding for putative effectors. The proportion of introduced adaptive variation and the potential for further introduction are a key focus of the present work.

Results & Discussion

Population genetic diversity across the genome of the emerging invasive *H. fraxineus* pathogen measured

among 43 isolates from across Europe is an eighth that of 15 isolates from a single wood in Japan. This general reduction in polymorphism caused by a bottleneck of two divergent individuals could reflect a reduction in the pathogens adaptive potential in its introduced environment. This reduction in genetic diversity is present in effectors as well as all other genes. Effectors in Europe retain a signal adaptive variation but far less than that present in the native range and we discuss the implications for the predicted level of virulence in Europe over the long term.

***Hymenoscyphus fraxineus* version2 genome approaches chromosome level contiguity**

The *Hymenoscyphus fraxineus* version 2 genome assembly (*Hf*-v2.0) is 62.08Mbp (Fig. 1). *Hf*-v2 contains the information present in 1672 contigs of *Hf*-v1.1 in just 23 scaffolds (Supplementary Information 2). Both *Hf*-v1.1 (published as a pre-print²⁹ and *Hf*-v2 were released (open) online and hosted on the OADB github repository (<https://github.com/ash-dieback-crowdsourcing>³⁰). The *Hf*-v2.0 gene annotation pipeline built on that of Saunders *et al.*²⁹ and identified 10,945 genes (11,097 transcripts, 1516.50 mean CDS length, 3.49 mean exons per gene; Supplementary Information 3). Telomeric repeats are identified at both ends of 14 scaffolds and, at one end of all but one of the remaining scaffolds (Supplementary Table 2). Close linkage between three pairs of scaffolds missing a telomere was detected in the offspring of an *H. fraxineus* cross using SNPs at the ends of scaffolds. After joining these scaffolds, we had an estimate of 20 chromosomes (pseudomolecules) for *H. fraxineus* (Fig. 1 track 2).

The European invasion population is bottlenecked and sexually recombining diversity from Asia

In 43 European individuals and 15 Japanese, we identified 6.26 million SNPs overall (SNPs in Japan (Jp) = 4.5×10^6 ; SNPs in Europe (EU) = 0.67×10^6 ; Supplementary information 5). A network of all genes shows us that the European population is genetically divergent and bottlenecked from that of the native population (Fig. 2). The European population is significantly smaller than the Japanese in terms of genetic diversity as the level of nucleotide diversity (π) of 43 individuals in Europe is an eighth of that observed in just 15 individuals in Japan (Wilcoxon (π), Jp genome > EU genome: $W(1257)=393210$, $p < 0.001$; Fig. 2). This

disparity in nucleotide diversity between Europe and Japan could, arguably, have been caused because the European population was founded from a population much smaller than that of Japan. Tajima's D^{31} is a statistic, centred around zero, that is sensitive to changes in effective population size (and/or the mode selection). In Japan, we observe a Tajima's D value close to zero across the genome ($\bar{x} = -0.22$) which indicates neutrality with purifying selection operating on genes which are more negative ($\bar{x} = -0.89$; Fig. 2). In Europe, the genomic signal is broadly positive ($\bar{x} = 1.28$; Fig. 2). This positive value is generated because the reduction in population size tends to remove high frequency alleles without equivalent reduction of rare alleles (Supplementary Fig. 10). This balancing of allele frequencies without a real reduction in SNP density by the loss of rare alleles across the genome is consistent with a bottlenecked founder EU population from a larger former population.

Sex is a fundamental determinant of plant pathogen adaptive potential not least because it allows recombination of host interaction genes, or effectors³². Here, as well as observing sexual reproduction between European isolates in the lab (Supplementary information 4) we see a breakdown in linkage disequilibrium in the wild in Europe (Supplementary Information 6). At the genome level Japanese and European populations are divergent ($F_{ST} = 0.58$ (95% CI = 0.54 - 0.56)) but recombination decouples relationships between genes. Across the genome, genetic differentiation correlates positively with diversity present in the Japanese population ($R^2 = 0.18$, $p < 0.001$) but negatively with diversity present in the European population ($R^2 = 0.51$, $p < 0.001$). That is to say that the more diversity present at a locus in Japan the less likely it is to be shared with Europe and, the more diversity present at a locus in Europe the more likely it is to be shared with Japan; a signal consistent with directionality from east to west (Fig. 1, tracks 6 & 7, Supplementary Fig. 11). Continued gene flow from the native range into Europe is important to prevent because sexual reproduction between divergent demes may be important for plant pathogen adaptation³³.

The European population was founded by two individuals from a large diverse population

Burokiene *et al.*²¹ reported similar levels of genetic diversity (based on 11 microsatellite loci) between central European populations and those on the invasion front. The authors suggest that for

introduced diversity to quickly spread to the invasion front either, the centre of genetic diversity quickly recombined and spread by range expansion, or only a small number of divergent *H. fraxineus* isolates arrived in Europe and present day diversity is the product of recombination among those divergent isolates. Here, we use third base positions of the coding regions of Core Eukaryotic Genes (CEGs) to distinguish between these scenarios and describe the invasion process. CEGs are highly conserved proteins essentially encoded in all eukaryotic genomes³⁴. Their importance to fundamental eukaryotic biology makes the 387 CEGs we identified in all *H. fraxineus* individuals an ideal set to explore the European invasion whilst minimising the impact of processes associated with polymorphism generation operating on other genes³⁵. The bottleneck into Europe has reduced the number of CEG haplotypes to 2.3 and 42% of all CEGs were monomorphic. Whereas, in Japan, the average number of haplotypes was 12.6 with no less than two haplotypes per locus. Strikingly, European CEGs are grouped into two divergent haplotypes and the majority of individuals have one or the other of the two main haplotypes. Furthermore, the level of divergence between those haplotypes reflects that across Japanese CEG networks (Fig. 3, Supplementary Fig. 18). It is our interpretation that these divergent haplotype pairs present in Europe have been introduced by two haploid founders as first suggested by Gross *et al.*³. Moreover, the divergence between these European haplotype pairs, without intermediates, represents the shadow of an ancestral population from which the European population was founded.

The observed estimate for the haploid effective population size ($\theta_\pi = 2N_e\mu$) in Japan is 2.46 million individuals ($\pi = 0.0246$) and the estimate from Europe is 0.67 million individuals ($\pi = 0.00672$; $\mu = 5 \times 10^{-9}$ per base per generation). However, this European estimate of nucleotide diversity is inflated by the divergence shadow of the two haploid individuals that founded the population and we used this divergent polymorphism to estimate the size of the source population from which this divergence shadow was generated. Coalescent simulations of a hypothetical European source of fixed effective population size which was bottlenecked to two (haploid) individuals shows us that an increase in effective population size, increases the average divergence between haplotypes in the two bottlenecked individuals. We find that an effective population size between 2.2 and 2.8 million individuals (closest median estimate = 2.5M) most

accurately replicates the observed diversity in present day European haplotypes (Fig. 4). Our estimate of the size of the source of the European *H. fraxineus* population is equivalent to that from a single Japanese wood. These estimates may reflect an expectation of the equilibrium diversity in any given *H. fraxineus* population but importantly also suggest that the European invaders could have come from a single site, perhaps even as a single ascocarp (fruiting body).

10% of *H. fraxineus* genes putatively interact with the host

The similarities between the effective population size of the Japanese and the estimated European-source populations suggest that the levels of genetic diversity observed in this native population reflect that of a representative European source. Effectors are a broad classification of proteins that interact with the host. They are secreted by bacteria, oomycetes, fungi, nematodes and aphids and they disable host defence components and facilitate colonisation. As such, effectors are in a co-evolutionary arms race with host resistance genes³⁶ and therefore can be studied for insights into pathogen adaptive potential. We used localisation signals of N-terminal presequences³⁷ to identify secreted proteins (i.e. putative effectors) throughout the genome of *H. fraxineus* (Fig. 1, track 5). From a starting set of 11,111 predicted proteins 1,132 are predicted effectors, which we clustered into 566 tribes (Supplementary Table 5). A search for N-terminal signal motifs found in putative effectors of other fungal pathogens did not reveal any conserved motif in *H. fraxineus* putative effectors. Density of AT base pairs are positively associated with repeat content ($R^2 = 0.79$, $p < 0.001$; Fig. 1, tracks 1&3; Supplementary Fig. 19). Both genes ($R^2 = 0.68$, $p < 0.001$; Fig. 1, track 4) and putative effectors ($R^2 = 0.08$, $p < 0.001$; Fig. 1, track 5) are negatively correlated with repeat regions and putative effectors tend to be found in regions of increased gene content ($R^2 = 0.11$, $p < 0.001$; Supplementary Fig. 20). Our observation that putative effector genes are spread across the genome similar to non-effector genes is consistent with the observation for the *Hf-v1.1* genome²⁹ and contrasts with that of the two-speed genome model, demonstrated in other filamentous plant pathogens³⁸. Perhaps the requirement for a dynamic process that operates to shuffle effector genes and generate novel effectors is lower in the genome of a sexually reproducing pathogen with high haplotype diversity³⁵. The

presence of sex and high haplotype diversity in turn indicate a long term balanced relationship between host and pathogen within the native range³⁹.

Structure and function of H. fraxineus putative effectors

Of the 1094 predicted and manually curated *H. fraxineus* effectors, 64% carry Pfam signatures suggesting highly diverse biological function (Supplementary Table 6). The largest subgroups include 127 proteins predicted to have glycosyl hydrolase activity, 77 have oxidoreductase activity, 71 carry a cytochrome P450, 56 harbour conserved domains of unknown function and 396 did not carry a Pfam domain. To our knowledge, *H. fraxineus* is the first sequenced fungal pathogen with such high number of predicted Cytochrome P450 proteins, and most of them possess putative secretion signals. Cytochrome P450s mode of action is typically via the monooxygenase reaction which is important in the generation and destruction of chemicals especially aromatics. In *H. fraxineus* they may be important in pathogenesis, especially as they carry a signal of diversifying selection (Supplementary information 7). Potential roles for P450s in pathogenesis include: 1) destruction of ash tissue-derived aromatic compounds with antifungal properties. Previous studies have demonstrated that plant secondary metabolism pathways are active during the infection process of *Magnaporthe oryzae*⁴⁰. 2) Penetration of ash tree tissues, P450 transcripts are upregulated in some pathogenic ascomycetes during the penetration of the plant cuticle, and are assumed to be used in the metabolism of hydrocarbon compounds which are the main constituents of host plants cuticle⁴¹. 3) Secreted *H. fraxineus* P450s could be used to facilitate fungal growth by extracellular oxidation of wood tissue, freeing nutrients, as well as enabling hyphal penetration^{42,43}. 4) P450s are part of the biosynthetic pathways for mycotoxins and phytohormones secreted by *H. fraxineus* during invasion of ash tissue, similar modes of actions have been shown for *Fusarium* multifunctional cytochrome P450 monooxygenase Tri4^{44,45}. Although most of the studied P450s in fungi are those with predicted intracellular activity, predicted secreted P450s in the ash dieback pathosystem could perform similar functions within specialised vesicles, the host apoplast or when transferred inside infected ash cells from invasive fungal structures. The presence of such high number of P450s predicted in the secretome of *H. fraxineus* suggests

biochemistry unique to *H. fraxineus*. Thus, these P450s are potential targets for novel fungicides with selective inhibitory action, similar to azole inhibitors that are selective for the fungal P450 gene CYP51, critical in sterol biosynthesis⁴⁶.

Transcribed low complexity domains may provide effectors with flexibility and diversity driven by adaptive evolution⁴⁷. Of the 1094 effector set, 23% contain short repeats (e.g. Ankyrin repeats, Tetratricopeptide repeats and Leucine-rich repeats; Supplementary Table 7). Some are predicted to have a nuclear localisation signal and are indicative of putative intracellular function in host cells while others are part of chitin recognition domains such as LysM, revealing possible apoplastic function. A large proportion of effectors (38%), are cysteine-rich with one to five predicted counts of structural disulfide bonds (Supplementary Table 8) which are predicted to confer stability, biological activity and resistance to proteases in the apoplastic space³⁷ and inside infected host cells⁴⁸.

Based on Gene ontology (GO) and BlastP search against fungal and oomycete proteins, we identify three effectors (one of them previously identified²⁹) with an NPP1 (necrosis-inducing Phytophthora protein) domain which is present in fungal, oomycete and bacterial proteins that induce hypersensitive reaction-like cell death upon infiltration into plant leaves⁴⁹. Two other proteins have predicted cell death activity and yet two more carry S1/P1 nuclease activity domains, which are involved in non-specific cleavage of RNA and single stranded DNA⁵⁰. Host cell death induction is suggested to be common among facultative parasites that engage in a necrotrophic lifestyle⁵¹.

Adaptive effector diversity is present in Europe, but is far greater in the native range

In sexually reproducing species genes may carry their own demographic history and here we can use estimates of nucleotide diversity, SNP effects and non-synonymous to synonymous polymorphism to understand the roles of these genes. We analysed the level of nucleotide diversity in all the genes, (including putative effectors) of the European and Japanese populations (Supplementary Table 9). First, there remains a significant positive correlation in the level of genetic diversity present in genes between the founder European and Japanese populations (gene π , $R^2 = 0.180$, $p < 0.001$; effector π , $R^2 = 0.144$, $p < 0.001$; Fig. 5),

despite a significant (73%) reduction in nucleotide diversity in Europe (Wilcoxon (π), Jp all genes > EU all genes: $W(21885) = 102060000$, $p < 0.001$; Fig. 6). This reduction in the level of nucleotide diversity has impacted genes (other than putative effectors) and putative effector genes alike as they are no different (Wilcoxon (π), EU genes ~ EU effectors: $W(10942) = 5420600$, $p = 0.78$; Fig. 6). Whereas, in Japan putative effector genes maintain significantly greater nucleotide diversity than that of other genes (Wilcoxon (π), Jp genes < Jp effectors: $W(10942) = 5671200$, $p < 0.041$; Fig. 6).

SNP effects can be characterised and we find that putative effectors in Europe have an increase in SNPs found to affect splice donor sites, or found in splice regions, and 5' UTRs (Supplementary information 7). Importantly however, there are also increases in SNP densities in synonymous, intron, upstream and downstream positions in putative effectors relative to other genes (Supplementary information 7). An increase in SNP density can be the product of mutation (with neutrality) and linkage to a locus under balancing selection (Hughes & Yeager 1998). Therefore, observed increases within these effector features could be the product of increased SNP density through linkage and balancing selection operating on effectors, rather than an indication of the importance of the feature *per se*. The overall pattern of SNP effects was similar in both the Japanese and European populations despite the overall reduction in diversity in Europe (Supplementary information 7).

Pathogens are under constant pressure to avoid detection by their hosts (van Valen 1974). Pathogen effectors must adapt to avoid recognition by the host surveillance system and are expected to undergo positive, diversifying, and balancing selection^{52,53}. To investigate the role of selection in the maintenance of polymorphism in putative effectors we assessed the level of adaptive diversity as measured using the mean pairwise ratio of non-synonymous (ka) to synonymous (ks) polymorphism present at all genes. The $ka:ks$ ratio is useful to detect the presence of adaptive evolution when applied regions of genes, such as binding domains, because a value greater than one is consistent with positive or diversifying selection (functional divergence). Moreover, a $ka:ks$ close to zero indicates negative or purifying selection. Here we apply the $ka:ks$ ratio across whole genes, which, we presume have different modes of selection operating across them. However, we expect that effectors will retain an overall higher $ka:ks$ value despite dilution by purifying

selection operating over the majority of the rest of the gene.

Approximately 97% of all *ka:ks* values in the European and Japanese effectors are lower than one.

Within this range, we observe a significant increase in the effector *ka:ks* over that of other genes between

$\omega = 0.2 - 0.5$ (Fig. 7). This represents a reduction in the strength of purifying selection in effectors relative

to other genes. At *ka:ks* values greater than one, we observe a significant peak in the density for effectors

which indicates a number of effectors evolving under contemporary positive selection (Fig. 7). Effectors in

Japan have higher *ka:ks* values than other genes (Wilcoxon (*ka:ks*), Jp genes < Jp effectors:

$W(10942) = 4891800$, $p < 0.001$; Fig. 8). In Europe, effector adaptive diversity has been maintained despite

the bottleneck, albeit at a lower level than in the native range (EU genes < EU effectors:

$W(10942) = 1750600$, $p = 0.045$; Fig. 8). Finally, consistent with linkage and balancing selection⁵⁴, the level

of synonymous polymorphism (*ks*) is also higher in effectors in both Japan and Europe (Wilcoxon (*ks*), Jp

genes < Jp effectors: $W(10942) = 4399800$, $p = 0.040$; Wilcoxon (*ks*), EU genes < EU effectors:

$W(10942) = 1804900$, $p = 0.001$; Fig. 8). We observe these *ka:ks* increases in effectors despite (presumed)

purifying selection operating across a portion of the coding region. Pseudogenes too, may have a *ka:ks* ratio

increased towards one. However, pseudogenisation could be more readily observed for effectors that are

increasingly recognised by the immune system⁵⁵. Here, we do not observe a significant increase in start lost,

stop gained or stop lost variants in effectors over other genes (Supplementary information 7, Table 10) but

more generally we believe that any impact of pseudogenisation on *ka:ks* is important to recognise.

Conclusions

The bottleneck of *H. fraxineus* into Europe removed the majority of neutral and much of the adaptive genetic

variation. Despite this, the pathogen has decimated ash from East to West. Host defence is characterised in

terms of levels of susceptibility⁵⁶ and with native effector (and genome) diversity high, further introduction

of pathogen genetic diversity runs the risk of increasing disease prevalence⁵⁷. Successive invasions can both

increase the level of genetic diversity above that of native populations and drive temporal fluctuations in

important pathogen diversity^{27,58}. Finally, an increased virulence at invasion onset, due to the density of

uninfected hosts^{59,60}, may be attenuated by natural selection on the pathogen as uninfected density reduces⁶¹ but this is much less probable where continued gene flow is allowed. For all of these reasons, it is most important to prevent further introduction of pathogen diversity to Europe.

Tree pathogens can be accidentally introduced because of commercial trade of live trees, wood and wood products¹ and we are unable to manage host resistance in the same way that we do for crops⁶². If we interpret the divergence between founder haploids to mean that multiple native genotypes can invade, we face a situation in which the majority of genes (91.9%) from a single additional haploid individual from the Japanese population would harbour a novel allele. Given the overall number of haplotypes present in Europe and Japan, any further migration from the East is likely to be accompanied by extreme increases to the level of adaptive polymorphism and disease severity of the European population.

Here we publish a high-quality *H. fraxineus* genome with annotation and we begin to analyse the population genetics of the European invasion using a limited sample of the native range. Further population genetics analyses including broader sampling of single isolates from the native range will allow robust modelling of the source meta-population and the invasion scenario. Moreover, population genetics of *de novo* assembled isolates can be used to unravel processes of genome evolution given the specifics of this two-haploid introduction. Important for the understanding of the disease progression is the combination of native effector nucleotide statistics with targeted RNA-seq experiments.

Notwithstanding the potential impact of an introduction of the emerald ash borer^{18,63}, current levels of European pathogen genetic diversity may infect 95% of all European ash. We must consider the implications of further introduction of diversity from the ash dieback pathogen.

Materials & Methods

Hymenoscyphus fraxineus collection and sequencing

A total of 44 *Hymenoscyphus fraxineus* isolates from Europe (21 UK, 13 Norway, 5 France, 4 Poland, 1 Austria) and nine from Japan (one haploid and eight fruiting bodies) were collected and cultured on malt extract agar¹³. Species confirmation was performed by assignment using PCR and sequencing of ITS

sequences. DNA was sequenced at either Edinburgh Genomics or The Earlham Institute on MiSeq and HiSeq Illumina platforms to between 25-160 sequence depth (Supplementary Table 1).

Genome Assembly and annotation of H. fraxineus

As part of a commitment to open science and rapid community analyses, the genome of isolate KW1, *Hf*-v1.1, was released online in March 2013 on the open ash dieback (oadb.tsl.ac.uk) repository²⁹. *Hf*-v2, published here and released online in February 2015, was assembled using a 200bp insert Paired End (PE) and a Long (5Kbp) Mate Pair (LMP) library. Overlapping PE reads were merged together using FLASH⁶⁴ and LMP reads were processed using FLASH and NextClip⁶⁵ with modifications as described in⁶⁶. SOAPec error correction (k=25, low coverage cut-off: PE = 4, LMP = 7) and SOAPDENOV0 v2.04⁶⁷ (pregraph -K 111 -p 200 -R; contig -M 1 -R -p 200; map; scaff) were used to assemble and scaffold (Supplementary Information 2).

The annotation pipeline used for genome *Hf*-v1.1²⁹ was replicated on *Hf*-v2 adding newer RNA-seq data available via the Open Ash Dieback repository (Supplementary Information 3). The *Hf*-v2 genome was repeat masked using *de novo* modelled repeats, known fungal and low complexity repeats (REPEATMODELER v1.0.8, REPEATMASKER v4.0.5⁶⁸). Annotations (9,737) from *Hf*-v1.1 were transferred to *Hf*-v2 using GMAP⁶⁹. AUGUSTUS v2.7⁷⁰ was used to predict additional protein-coding genes using protein and RNA-seq alignments (16 libraries, Supplementary Table 16) with the previously trained *Hymenoscyphus fraxineus* model. We ran a transcript extension protocol on all genes to identify whether those genes within 100bp of an upstream gene had evidence for extension. In addition, we checked this extended gene dataset for potential secreted proteins. This protocol allowed the identification of 36 genes that had not previously been recognised as having a methionine start codon which were included into our effector-mining pipeline.

We used the pipeline described by Saunders et al.^{29,37} to identify putative effectors. Briefly, SIGNALP4 (-t euk -s notm -u 0.34) was run to identify transcripts with a signal peptide and from these TMHMM and TARGETP were used to remove transcripts with transmembrane or mitochondrial localisation signals⁷¹. Finally, secreted proteins were clustered into tribes using TRIBEMCL v12.135 (e-value 10⁻⁵)⁷².

We used T-REKS⁷³ to identify repeats, the Disulfide algorithm⁷⁴ to predict disulfide bonds and PREDICTNLS⁷⁵ to predict nuclear localisation signals.

Mapping and SNP calling

Reads were trimmed for Illumina adapters and Phred quality (-q20) and discarded where length was reduced below 80bp (TRIM GALORE v0.3.3; Babraham Institute, Cambridgeshire, UK). Read alignment and mapping were performed using BWA-MEM v0.7.7⁷⁶. SAMTOOLS v0.1.19 and BCFTOOLS v0.1.19-44428cd⁷⁷ were used to sort, remove duplicate reads, mpileup (-D) and call variants (bcftools view -cg). VCFTOOLS v0.1.13⁷⁸ was used to filter variants to a minimum depth of ten, maximum depth of $1.8 \times$ mean individual depth and a SNP genotype quality of at least 30. SNP sites with more than two alleles were excluded as likely errors, as were sites reported as heterozygous in haploids. Finally, sites that were missing in 20% or more individuals were also removed. Three samples were removed from further analyses because of insufficient depth or evidence of contamination (20 UK, 13 Norway, 5 France, 4 Poland, 1 Austria, 8 Japan (1 haploid 7 fruiting bodies); see Supplementary Information 1).

Phasing fruiting body samples

SHAPEIT2⁷⁹, used to phase fruiting body data, first applies a phase informative read step (extractPIRs) to group SNPs present on the same read pair and then phases remaining SNPs using population level polymorphism in the second step (assemble --states 1000 --burn 60 --prune 60 --main 300 --effective-size 564000 --window 0.5). All Japanese samples including the haploid strain were used in the phasing step. As part of the phasing process indels must be removed from the dataset. Effective population size ($\theta = 2N_e\mu$) was calculated assuming a mutation rate of 5×10^{-9} per generation per site⁸⁰ using the mean nucleotide diversity present in 100Kbp windows in the Japanese population.

SNP diversity and divergence analysis

Polymorphism statistics and sliding windows i.e. sites per individual, SNP depth, transition/transversion (Ts/Tv) ratio, linkage (--thin 1000; --ld-window-bp 500000), π and F_{ST} were conducted using VCFTOOLS v0.1.13 & BEDTOOLS v2.22.0⁸¹ & custom scripts (<https://github.com/mcmullan0>). SNPEFF V4.2⁸² was used to categorise the impact of variants. Effectors are, on average, shorter than other genes and so we controlled

for the number of variants present per feature by dividing by feature length. DNASP v5.10.01⁸³ was used to calculate population genetic statistics per gene after fasta conversion using GATK v3.5.0 FastaAlternateReferenceMaker⁸⁴ and PAML v4.9⁸⁵ was used (YN00) in pairwise mode to calculate the average *ka:ks* ratio nonsynonymous (*ka*) and synonymous (*ks*) values for each gene with at least one synonymous mutation. The sm package⁸⁶ was used in R v3.2.1 to compare the density distributions of *ka:ks*. Output data for each gene (including effectors) from SNPEFF, DNASP & PAML are combined into a single table. Data are represented using SPLITSTREE v4.13.1⁸⁷ using default parameters (CDS of all genes and separately for each CEG) and CIRCOS v0.67-5⁸⁸.

Parental cross and linkage analysis

Two *H. fraxineus* isolates of opposite mating types, LWD054 and LWD067, were collected from stem lesions on trees in Norfolk, UK in 2014. Crosses were made between parental lines according to the method of Wey *et al.*⁸⁹ except that isolates were kept at 25°C in 16 hours light. KASPTM⁹⁰ markers were designed based on SNP sites between two *H. fraxineus* parents mapped to the *Hf*-v2 assembly (Supplementary Information 4). Linkage between scaffolds was ascertained using Chi-squared test for linkage of sites close to ends of the 23 scaffolds, across 28 F1 offspring. Scaffolds at linkage group ends were examined for the presence of telomeric motifs.

Statistics

R v3.2.1 was used to fit correlations to test the level of association between population signals in order to show directionality in the invasion and portray the bottleneck behind Tajima's D. In addition, we correlated feature coverage (i.e. repeat content) of regions with gene and effector coverage to understand genome evolution. Chi-squared tests were used to assess linkage between scaffolds in the offspring of a parental cross. Two-sample Wilcoxon tests (one tailed indicated by '>' or '<'; two tailed indicated by '~'), were used to test for differences in signals of nucleotide diversity, Tajima's D and *ka:ks* between non-effector genes and effectors in European and Japanese populations. Bootstrap analyses were used to compare variants effect by feature or gene (e.g. SnpEff output) and involved sampling feature with replacement (×1000). Bootstrapped comparisons are tested using a randomisation test and 95% confidence intervals (CI) are

407 presented.

408 *Ancestral effective population size*

409 In order to estimate the size of the population from which the European population was founded we used the
 410 SNPs within the third base pair of codons within 387 core eukaryotic genes. CEGMA v2.4³⁴ initially
 411 identified 440 CEGs which was reduced to 387 genes that were confirmed by reciprocal blastp (BLAST
 412 v2.2.31⁹¹ (-evalue 1e-5) with a minimum of 90% coverage (to remove spurious technical errors in the
 413 assembly and/or biological duplicates), removal of CEGs with nonsense variants (expected pseudogenes)
 414 and a minimum mean depth of 10x per gene across all individuals (remove poorly covered CEGs). Each
 415 gene was then run through a pipeline to quantify the pairwise distance between each allele (using
 416 SPLITSTREE) and the diversity per gene (using DNASP v5).

417 The average number of base pairs (3rd codon bp) per CEG was 448 and this number was used as an
 418 input to FASTSIMCOAL2 v2.5.2.8⁹², a fast sequential markov coalescent simulator used to estimate the size of
 419 the population from which the founders into Europe came. A simple model (x1000) of a single (haploid)
 420 population of fixed effective size ($N_e = 100,000 - 4,000,000$) containing 387 unlinked CEGs that were freely
 421 mutating at a rate of 5×10^{-9} per base per generation⁸⁰ was bottlenecked to two individuals. From these two
 422 individuals, we recorded the number of haplotypes and the haplotype divergence at each gene for
 423 comparison to observed data.

424

425

426

427

428

Data access

All MP and LMP sequencing data generated for the *Hf*-v2.0 genome (PRJEB21027), population genetics and parental cross reads have been submitted to the European Nucleotide Archive (under projects PRJEB21061, PRJEB21062, PRJEB21063, PRJEB21064, PRJEB21059, PRJEB21060; see Supplementary Table 1 for accession Nos). The genome annotation is available at the Earlham Institute Open Data site (http://opendata.earlham.ac.uk/Hymenoscyphus_fraxineus/EI/v2/).

Author contributions

MM, CF, JB, MB, AD & MC contributed to the design of the project. EO, AE, KY, TH, LW, PJ, RI, CH, AH, AV, HS & AD contributed to collection of samples. Acquisition and processing of sequence data was done by MM, WV, AE, JW, GeoK & DM. MM, MR, GemK, BC, GC, LB, EO, DGS, LPA, BW, DS, JB, & MC contributed to the analysis and interpretation of data. MM, MR, GK, BC, LPA, LB, EO, AH, NH DS JB AD & MC contributed to the drafting and revising of the MS.

Acknowledgments

We thank The Sainsbury Laboratory for hosting the Open Ash Dieback Git repository (oadb.tsl.ac.uk), and Sophie Janacek for assistance with project management. The submission of sequencing data was brokered by COPO (<http://copo-project.org/>). The Ash Dieback research project was funded jointly by the Biotechnology and Biological Sciences Research Council (BBSRC) and the Department of Environment Food and Rural Affairs (Defra) via the ‘Normex’ project (BBS/E/J/000CA523) and Economic and Social Research Council, the Forestry Commission, the Natural Environment Research Council (NERC) and the Scottish Government, under the Tree Health and Plant Biosecurity Initiative (BB/L01291X/1). The UMR IAM is supported by a grant overseen by the French National Research Agency (ANR) as part of the “Investissements d’Avenir” program (ANR-11-LABX-0002-01, Laboratory of Excellence ARBRE). Genome sequencing was done by Edinburgh Genomics, and at the Earlham Institute (EI, formerly The Genome Analysis Centre, Norwich), by members of the Platforms and Pipelines Group funded by BBSRC National Capability in Genomics grant (BB/J010375/1). Additional research funding was via a BBSRC Institute

456 Strategic Programme Grant for Bioinformatics (BB/J004669/1).

457

Figures

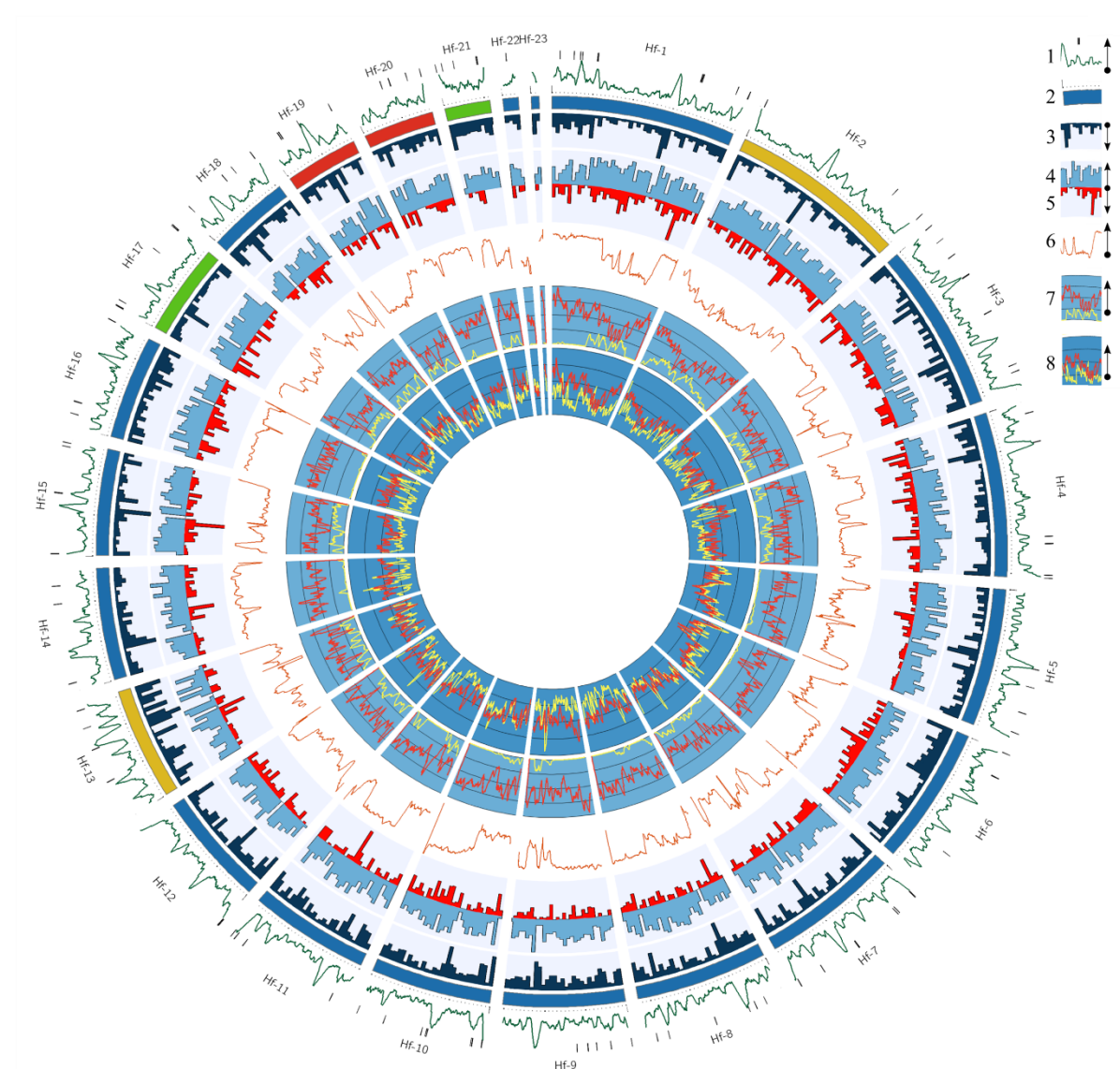


Figure 1. *H. fraxineus* genome organisation (repeats, effectors etc.) and population statistics (F_{ST} , π etc.). (1) AT richness sliding window (63Kbp-5Kbp) with ticks indicating regions of n's or gaps in the assembly >2.5Kbp. (2) Scaffolds (ticks = 100Kbp) with evidence for linkage between red, green and yellow pairs of scaffolds. (3) Repeat density (100Kbp). (4 & 5) Gene density & effector density, respectively (100Kbp). (6) Genetic differentiation (F_{ST}) sliding window between Japanese and European populations (100Kbp-5Kbp). (7) Nucleotide diversity (π) sliding window within Japanese (red) and European (yellow) populations (100Kbp-5Kbp). (8) SNP density sliding window within Japanese (red) and European (yellow) populations (100Kbp-5Kbp).

populations (100Kbp-5Kbp). Legend shows track number and orientation.

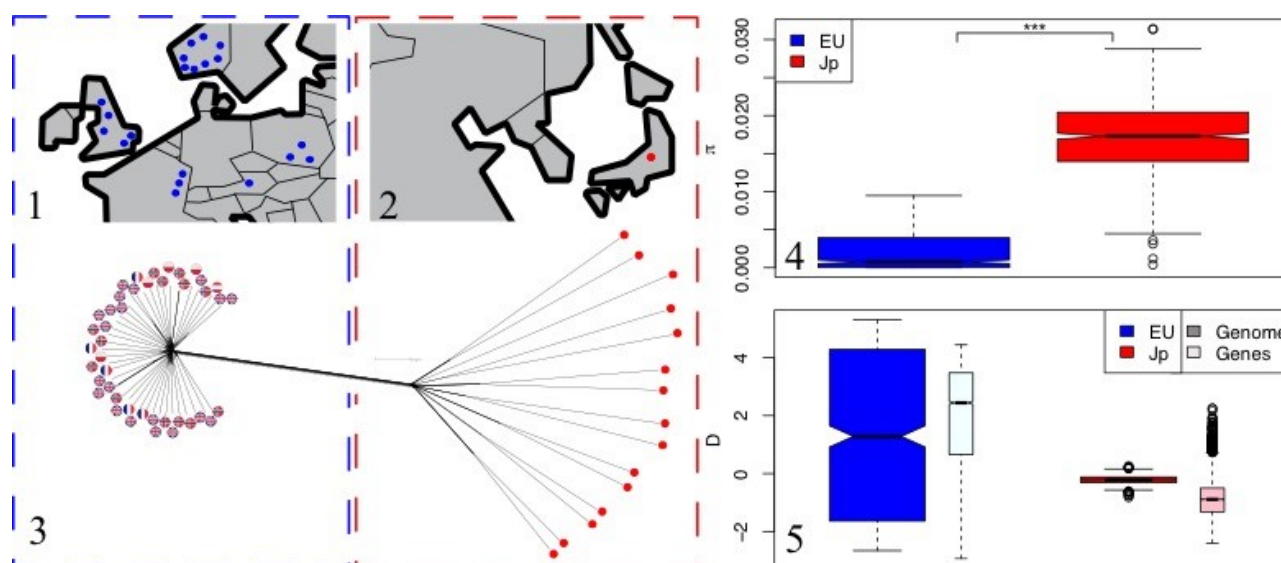


Figure 2. *H. fraxineus* European and Japanese sample sites and population genetic diversity. (1 & 2) European and Japanese sampling locations. (3) A network of the coding regions of all genes in the genome (16,589,637bp) bounded by dotted line indicating the sampling location in panels 1 & 2. (4) Nucleotide diversity (π 100Kbp windows) in both Japanese and European populations is greater in Japan than in Europe. (5) Tajima's D for all SNPs (100Kbp windows) across the genome and again in all genes within the Japanese and European populations. In Japan, a signal of neutrality is present across the genome and increases in the genes (purifying selection). The European population has a much broader positive distribution, consistent with a recent population bottleneck.

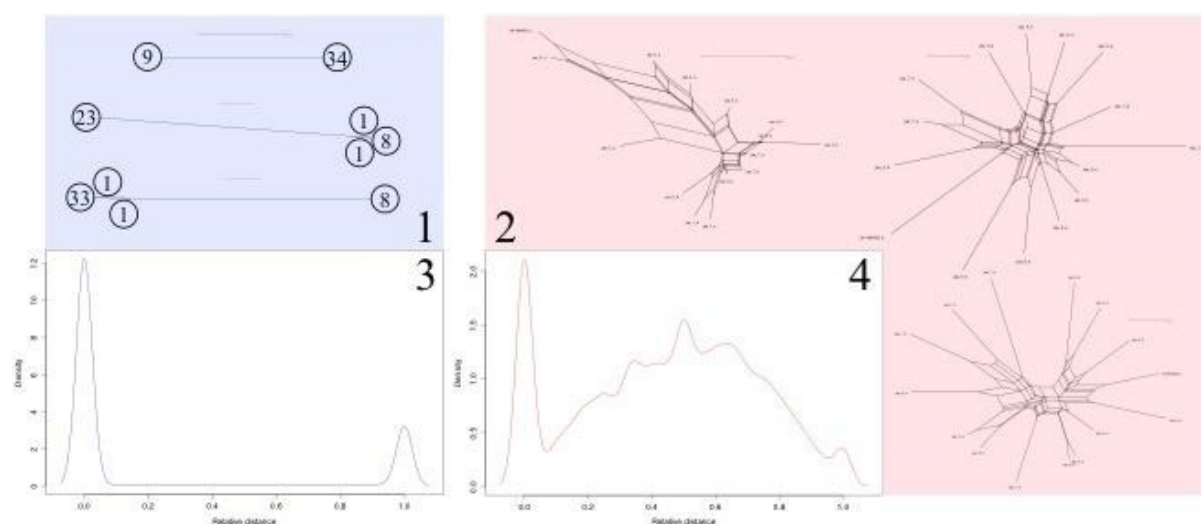


Figure 3. Divergence between haplotypes in Europe is high and bimodal. Networks were generated using the third base position of 387 CEGs. Three CEG networks are shown from European (1) and Japanese *H. fraxineus* (2) populations. Density plots show the relative pairwise distance amongst haplotypes of 387 CEG in Europe (3) and Japan (4). In Japan, pairwise distances range in their divergence. In Europe, the majority of genes are either identical or at the complete opposite ends of the divergence distribution (numbered in circle). This represents the presence of two major divergent haplotype groups (Supplementary Fig. 18).

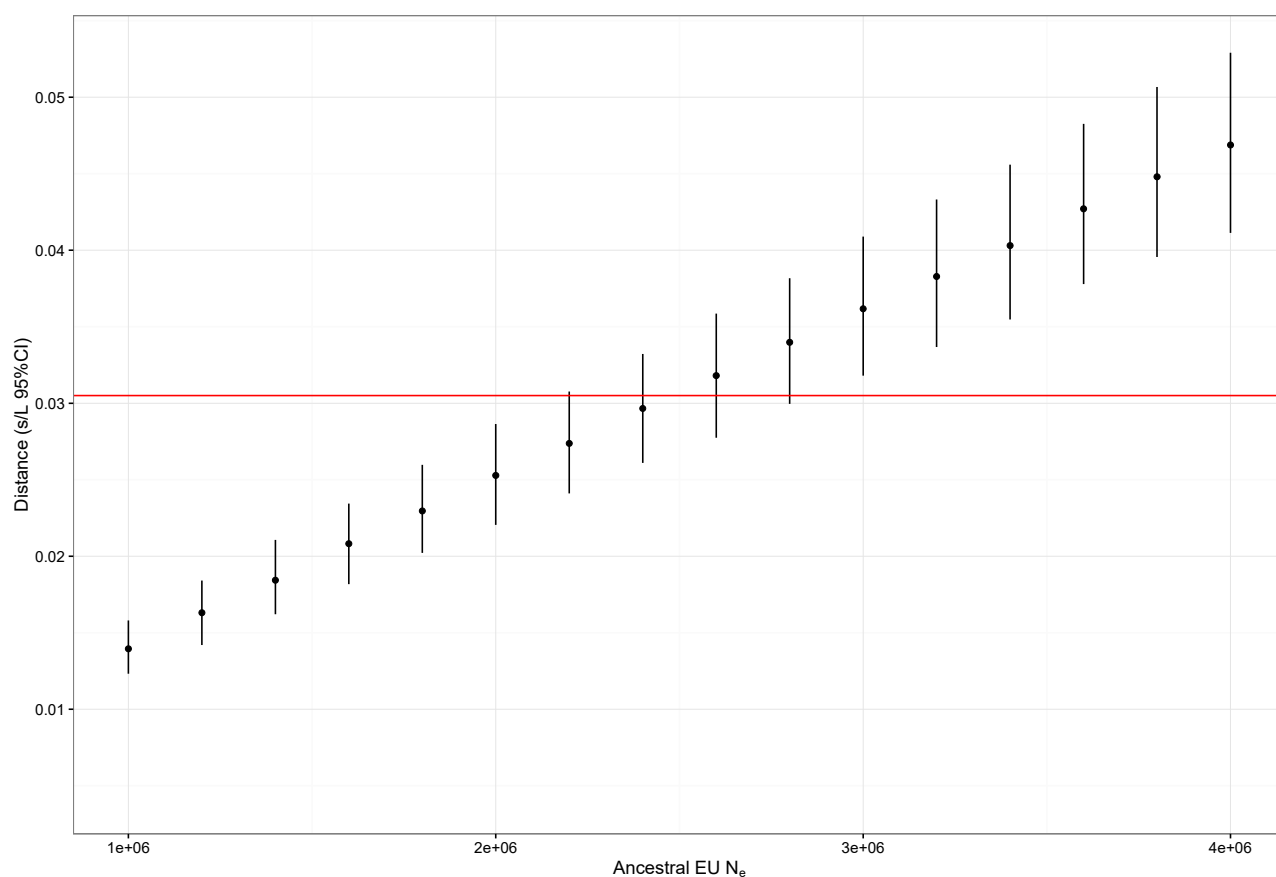


Figure 4. Coalescent simulations show a population of 2.5 million individuals maintain enough polymorphism to account for the observed haplotype divergence. The coalescent of a population was simulated (x 1000) at an effective population size of between 1M and 4M haploid individuals. At equilibrium, this population was bottlenecked to two individuals and the divergence at all loci with one or more polymorphisms was recorded. 2.5M individuals best encompassed the observed diversity in present day shared divergent haplotypes (observed SNPs/bp = 0.0305).

499



500

501 **Figure 5. The nucleotide diversity (π) in all genes in the Japanese and European populations is**
 502 **positively correlated.** This association demonstrates the association of polymorphism present in genes in
 503 Japan and Europe even after a population bottleneck. This reduction in nucleotide diversity for both genes
 504 and putative effectors in Europe is not significantly different between these groups of genes.

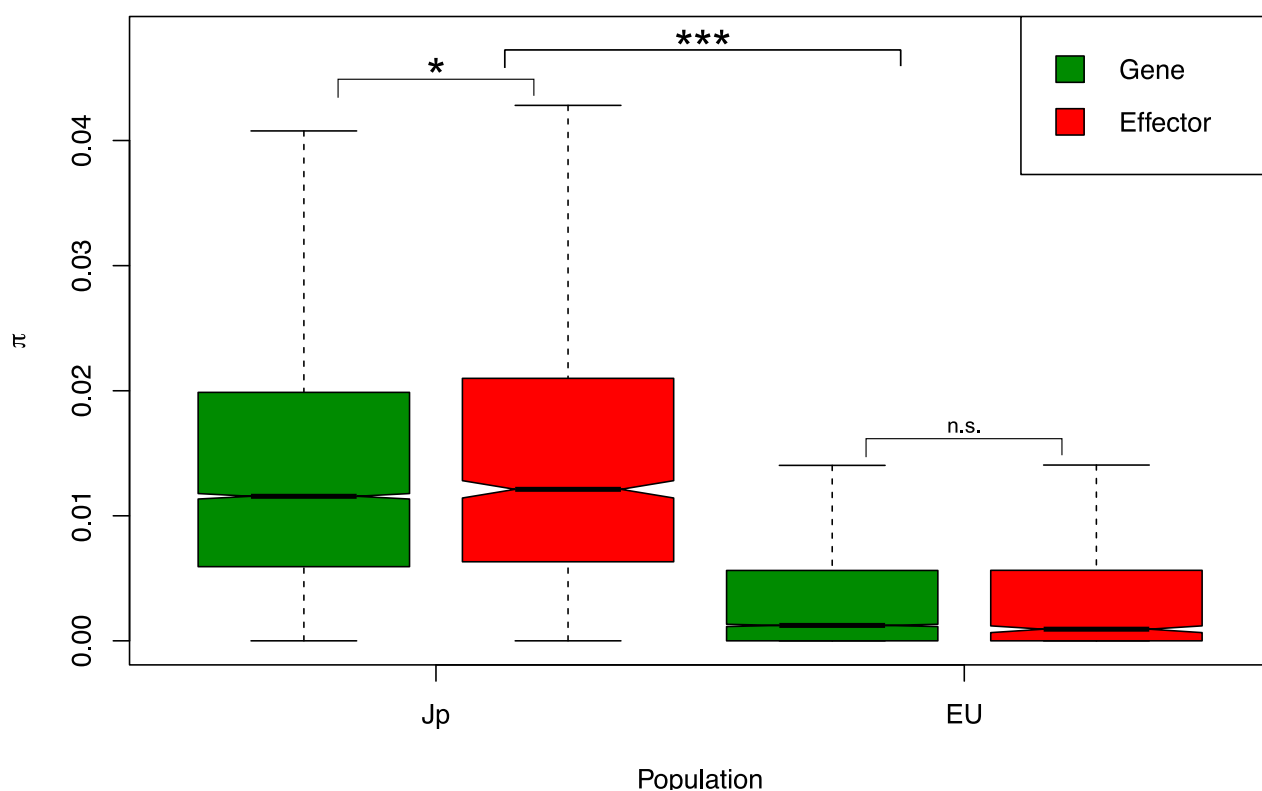


Figure 6. Nucleotide diversity (π) present at putative effectors and all other genes within Japanese and European populations. Nucleotide diversity (outliers not shown) over all genes combined is significantly greater in Japan (Jp) than in the Europe (EU). Effectors have significantly greater nucleotide diversity than all other genes in Japan but not in Europe.

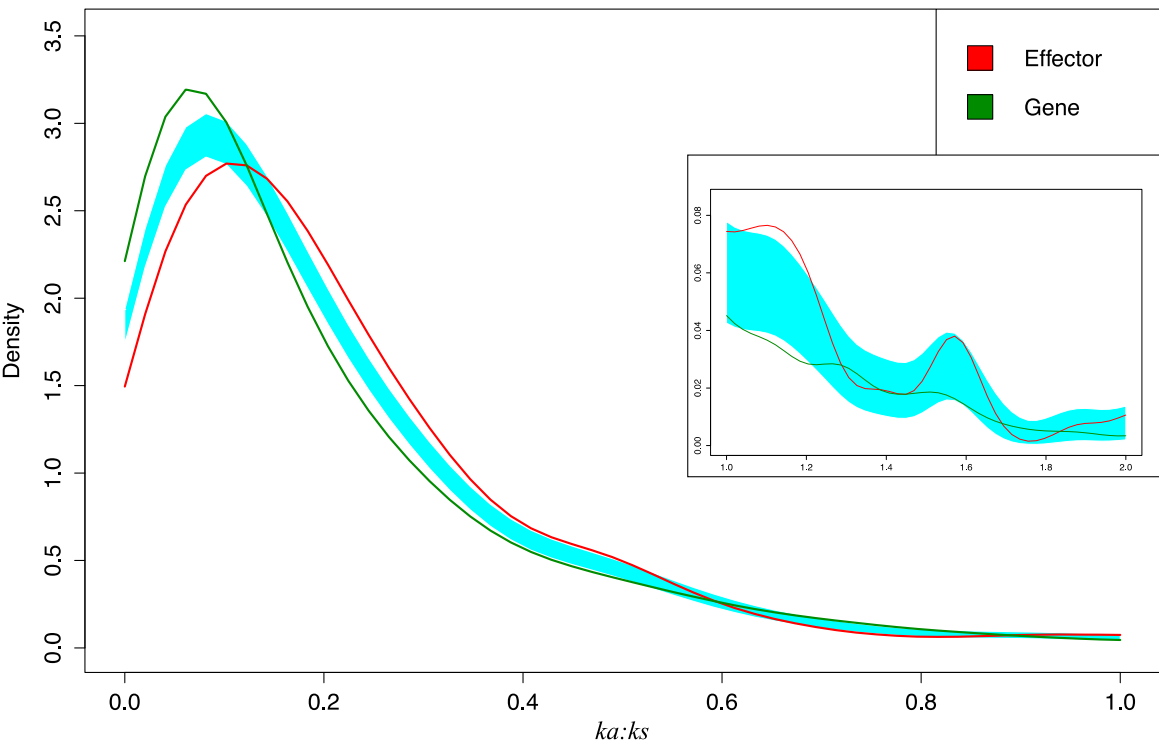


Figure 7. $ka:ks$ density for genes and effectors in European and Japanese populations combined. $ka:ks$ is significantly higher for effectors over a considerable range of the distribution between zero and one ($p < 0.001$). This is evidence for reduced efficacy of purifying selection in these genes. $ka:ks$ greater than one (less than two; inset) also contains a peak for effector genes which is consistent with the operation of positive selection driving adaptive change in these effectors ($p = 0.01$).

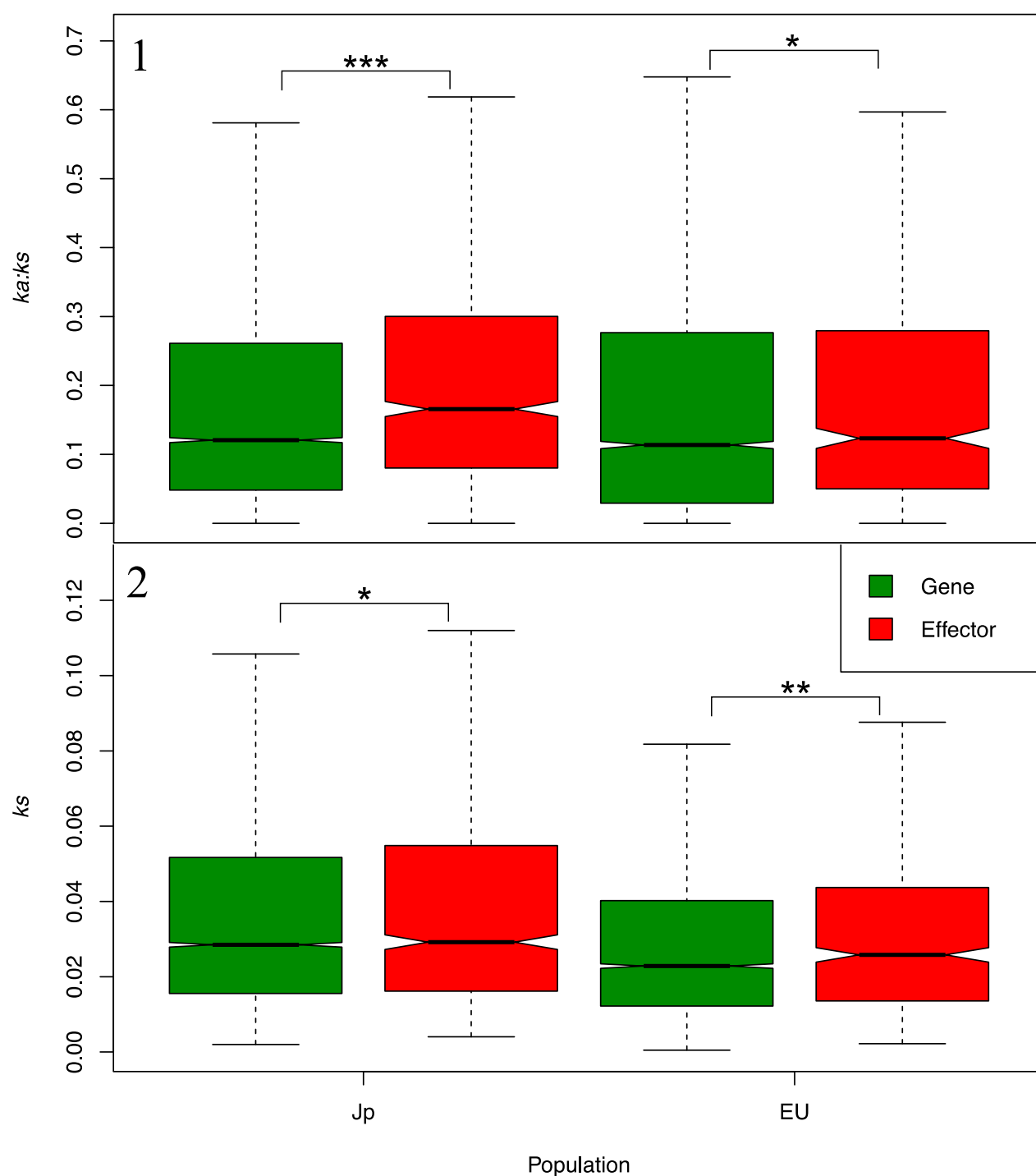


Figure 8. $ka:ks$ and synonymous site density (ks) for effectors and all other genes. Outliers not shown. The signal of polymorphism visible to natural selection is greater in effectors than other genes in both Japan and Europe. However, the strength of this signal has been reduced in the European population. Synonymous site density is significantly higher in effectors in both the European and Japanese populations.

526

527 References

- 528 1. Ghelardini, L., Pepori, A. L., Luchi, N., Capretti, P. & Santini, A. Drivers of emerging fungal
529 diseases of forest trees. *For. Ecol. Manage.* **381**, 235–246 (2016).
- 530 2. Brasier, C. M. & Buck, K. W. Rapid evolutionary changes in a globally invading fungal pathogen
531 (Dutch elm disease). *Biol. Invasions* **3**, 223–233 (2001).
- 532 3. Gross, A., Hosoya, T. & Queloz, V. Population structure of the invasive forest pathogen
533 *Hymenoscyphus pseudoalbidus*. *Mol. Ecol.* **23**, 2943–2960 (2014).
- 534 4. Pautasso, M., Aas, G., Queloz, V. & Holdenrieder, O. European ash (*Fraxinus excelsior*) dieback - A
535 conservation biology challenge. *Biol. Conserv.* **158**, 37–49 (2013).
- 536 5. Drenkhan, R. *et al.* *Hymenoscyphus fraxineus* is a leaf pathogen of local *Fraxinus* species in the
537 Russian Far East. *Plant Pathol.* **66**, 490–500 (2017).
- 538 6. Gross, A. & Holdenrieder, O. Pathogenicity of *Hymenoscyphus fraxineus* and *Hymenoscyphus albidus*
539 towards *Fraxinus mandshurica* var: *Japonica*. *For. Pathol.* **45**, 172–174 (2015).
- 540 7. Cleary, M. *et al.* Friend or foe? Biological and ecological traits of the European ash dieback pathogen
541 *Hymenoscyphus fraxineus* in its native environment. *Sci. Rep.* **6**, 21895 (2016).
- 542 8. Gilbert, G. S. & Webb, C. O. Phylogenetic signal in plant pathogen-host range. *Proc. Natl. Acad. Sci.*
543 **104**, 4979–4983 (2007).
- 544 9. Roy, B. A. *et al.* Increasing forest loss worldwide from invasive pests requires new trade regulations.
545 *Front. Ecol. Environ.* **12**, 457–465 (2014).
- 546 10. Nielsen, L. R., McKinney, L. V., Hietala, A. M. & Kj?r, E. D. The susceptibility of Asian, European
547 and North American *Fraxinus* species to the ash dieback pathogen *Hymenoscyphus fraxineus* reflects
548 their phylogenetic history. *Eur. J. For. Res.* **136**, 59–73 (2017).
- 549 11. Blackburn, T. M., Lockwood, J. L. & Cassey, P. The influence of numbers on invasion success. *Mol.*
550 *Ecol.* n/a-n/a (2015). doi:10.1111/mec.13075
- 551 12. Cross, H. *et al.* Fungal diversity and seasonal succession in ash leaves infected by the invasive

- 552 ascomycete *Hymenoscyphus fraxineus*. *New Phytol.* 1405–1417 (2017). doi:10.1111/nph.14204
- 553 13. Kowalski, T. *Chalara fraxinea* sp. nov. associated with dieback of ash (*Fraxinus excelsior*) in Poland.
554 *For. Pathol.* **36**, 264–270 (2006).
- 555 14. Zhao, Y., Hosoya, T. & Baral, H. *Hymenoscyphus pseudoalbidus*, the correct name for *Lambertella*
556 *albida* reported from Japan. *Mycotaxon* **122**, 25–41 (2012).
- 557 15. Mckinney, L. V. *et al.* The ash dieback crisis: Genetic variation in resistance can prove a long-term
558 solution. *Plant Pathol.* **63**, 485–499 (2014).
- 559 16. Harper, A. L. *et al.* Molecular markers for tolerance of European ash (*Fraxinus excelsior*) to dieback
560 disease identified using Associative Transcriptomics. *Sci. Rep.* **6**, (2016).
- 561 17. Mitchell, R. J. *et al.* *The potential ecological impact of ash dieback in the UK. The potential*
562 *ecological impact of ash dieback in the UK* (2014).
- 563 18. Thomas, P. A. Biological Flora of the British Isles: *Fraxinus excelsior*. *J. Ecol.* **104**, 1158–1209
564 (2016).
- 565 19. Gross, A., Zaffarano, P. L., Duo, A. & Grünig, C. R. Reproductive mode and life cycle of the ash
566 dieback pathogen *Hymenoscyphus pseudoalbidus*. *Fungal Genet. Biol.* **49**, 977–86 (2012).
- 567 20. Fones, H. N., Mardon, C. & Gurr, S. J. A role for the asexual spores in infection of *Fraxinus excelsior*
568 by the ash-dieback fungus *Hymenoscyphus fraxineus*. *Sci. Rep.* **6**, 34638 (2016).
- 569 21. Burokiene, D. *et al.* Genetic population structure of the invasive ash dieback pathogen
570 *Hymenoscyphus fraxineus* in its expanding range. *Biol. Invasions* **17**, 2743–2756 (2015).
- 571 22. McDonald, B. A. & Linde, C. Pathogen population genetics, evolutionary potential, and durable
572 resistance. *Annu. Rev. Phytopathol.* **40**, 349–379 (2002).
- 573 23. Charlesworth, B. Effective population size and patterns of molecular evolution and variation. *Nat.*
574 *Rev. Genet.* **10**, 195–205 (2009).
- 575 24. Estoup, A. *et al.* Is There A Genetic Paradox of Biological Invasion? *Annu. Rev. Ecol. Evol. Syst.* **47**,
576 51–72 (2016).
- 577 25. Allendorf, F. W. & Lundquist, L. L. Introduction: Population Biology, Evolution, and Control of

- 578 Invasive Species. *Conserv. Biol.* **17**, 24–30 (2003).
- 579 26. Goodwin, S. B., Cohen, B. A. & Fry, W. E. Panglobal distribution of a single clonal lineage of the
580 Irish potato famine fungus. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 11591–5 (1994).
- 581 27. Uller, T. & Leimu, R. Founder events predict changes in genetic diversity during human-mediated
582 range expansions. *Glob. Chang. Biol.* **17**, 3478–3485 (2011).
- 583 28. Fry, W. E. *Phytophthora infestans* : New Tools (and Old Ones) Lead to New Understanding and
584 Precision Management. *Annu. Rev. Phytopathol.* **54**, 529–547 (2016).
- 585 29. Saunders, D. *et al.* Crowdsourced analysis of ash and ash dieback through the Open Ash Dieback
586 project: A year 1 report on datasets and analyses contributed by a self-organising community. *bioRxiv*
587 (2014). doi:<https://doi.org/10.1101/004564>
- 588 30. Maclean, D. *et al.* Crowdsourcing genomic analyses of ash and ash dieback - power to the people.
589 *Gigascience* **2**, 2 (2013).
- 590 31. Tajima, F. Statistical-method for testing the neutral mutation hypothesis by DNA polymorphism.
591 *Genetics* **123**, 585–595 (1989).
- 592 32. Croll, D., Lendenmann, M. H., Stewart, E. & McDonald, B. A. The Impact of Recombination
593 Hotspots on Genome Evolution of a Fungal Plant Pathogen. *Genetics* **201**, 1213–1228 (2015).
- 594 33. McMullan, M. *et al.* Evidence for suppression of immunity as a driver for genomic introgressions and
595 host range expansion in races of *Albugo candida*, a generalist parasite. *Elife* **4**, 1–24 (2015).
- 596 34. Parra, G., Bradnam, K. & Korf, I. CEGMA: a pipeline to accurately annotate core genes in eukaryotic
597 genomes. *Bioinformatics* **23**, 1061–7 (2007).
- 598 35. Stukenbrock, E. H. & Croll, D. The evolving fungal genome. *Fungal Biol. Rev.* **28**, 1–12 (2014).
- 599 36. Dangl, J. L., Horvath, D. M. & Staskawicz, B. J. Pivoting the Plant Immune System from Dissection
600 to Deployment. *Science (80-.).* **341**, 746–51 (2013).
- 601 37. Saunders, D. G. O. *et al.* Using hierarchical clustering of secreted protein families to classify and rank
602 candidate effectors of rust fungi. *PLoS One* **7**, (2012).
- 603 38. Raffaele, S. & Kamoun, S. Genome evolution in filamentous plant pathogens: why bigger can be

better. *Nat Rev Micro* **10**, 417–430 (2012).

39. Tellier, A. & Brown, J. K. . Stability of genetic polymorphism in host-parasite interactions. *Proc. R. Soc. B Biol. Sci.* **274**, 809–817 (2007).

40. Collemare, J. *et al.* Magnaporthe grisea avirulence gene ACE1 belongs to an infection-specific gene cluster involved in secondary metabolism. *New Phytol.* **179**, 196–208 (2008).

41. Oh, Y. *et al.* Transcriptome analysis reveals new insight into appressorium formation and function in the rice blast fungus Magnaporthe oryzae. *Genome Biol.* **9**, R85 (2008).

42. Shary, S. *et al.* Differential expression in Phanerochaete chrysosporium of membrane-associated proteins relevant to lignin degradation. *Appl. Environ. Microbiol.* **74**, 7252–7257 (2008).

43. K, S. & JS, Y. P450 monooxygenases (P450ome) of the model white rot fungus Phanerochaete chrysosporium. *Crit. Rev. Microbiol.* **38**, 339–363 (2012).

44. Tudzynski, B., Rojas, M. C., Gaskin, P. & Hedden, P. The gibberellin 20-oxidase of Gibberella fujikuroi is a multifunctional monooxygenase. *J. Biol. Chem.* **277**, 21246–21253 (2002).

45. Tokai, T. *et al.* Fusarium Tri4 encodes a key multifunctional cytochrome P450 monooxygenase for four consecutive oxygenation steps in trichothecene biosynthesis. *Biochem. Biophys. Res. Commun.* **353**, 412–417 (2007).

46. Warrilow, A. G., Parker, J. E., Kelly, D. E. & Kelly, S. L. Azole Affinity of Sterol 14-Demethylase (CYP51) Enzymes from Candida albicans and Homo sapiens. *Antimicrob. Agents Chemother.* **57**, 1352–1360 (2013).

47. Mesarich, C. H., Bowen, J. K., Hamiaux, C. & Templeton, M. D. Repeat-containing protein effectors of plant-associated organisms. *Front. Plant Sci.* **6**, 872 (2015).

48. Blondeau, K. *et al.* Crystal structure of the effector AvrLm4-7 of Leptosphaeria maculans reveals insights into its translocation into plant cells and recognition by resistance proteins. *Plant J.* **83**, 610–624 (2015).

49. Fellbrich, G. *et al.* NPP1, a Phytophthora-associated trigger of plant defense in parsley and Arabidopsis. *Plant J.* **32**, 375–390 (2002).

- 630 50. Balabanova, L. A. *et al.* An Extracellular S1-Type Nuclease of Marine Fungus *Penicillium melinii*.
631 *Mar. Biotechnol.* **14**, 87–95 (2012).
- 632 51. Gijzen, M. & Nürnberger, T. Nep1-like proteins from plant pathogens: Recruitment and
633 diversification of the NPP1 domain across taxa. *Phytochemistry* **67**, 1800–1807 (2006).
- 634 52. Gladieux, P. *et al.* The population biology of fungal invasions. *Mol. Ecol.* **24**, 1969–1986 (2015).
- 635 53. Stukenbrock, E. H. & McDonald, B. a. Population genetics of fungal and oomycete effectors
636 involved in gene-for-gene interactions. *Mol. Plant. Microbe. Interact.* **22**, 371–80 (2009).
- 637 54. Hughes, A. L. & Yeager, M. Natural selection at major histocompatibility complex loci of
638 vertebrates. *Annu. Rev. Genet.* **32**, 415–435 (1998).
- 639 55. Burstein, D. *et al.* Genomic analysis of 38 *Legionella* species identifies large and diverse effector
640 repertoires. *Nat. Genet.* **48**, 167–175 (2016).
- 641 56. Sollars, E. S. *et al.* Genome sequence and diversity of European ash trees threatened by ash dieback.
642 *Nature* 212–216 (2017). doi:10.1038/nature20786
- 643 57. Gross, A. & Sieber, T. N. Virulence of *Hymenoscyphus albidus* and native and introduced
644 *Hymenoscyphus fraxineus* on *Fraxinus excelsior* and *Fraxinus pennsylvanica*. *Plant Pathol.* **65**, 655–
645 663 (2016).
- 646 58. McMullan, M. & van Oosterhout, C. Inference of selection based on temporal genetic differentiation
647 in the study of highly polymorphic multigene families. *PLoS One* **7**, e42119 (2012).
- 648 59. Read, A. F. The evolution of virulence. *Trends Microbiol.* **2**, 73–76 (1994).
- 649 60. Lenski, R. E. & May, R. M. The evolution of virulence in parasites and pathogens: reconciliation
650 between two competing hypotheses. *Journal of theoretical biology* **169**, 253–265 (1994).
- 651 61. Berngruber, T. W., Froissart, R., Choisy, M. & Gandon, S. Evolution of Virulence in Emerging
652 Epidemics. *PLoS Pathog.* **9**, (2013).
- 653 62. Brown, J. K. M. Durable Resistance of Crops to Disease: A Darwinian Perspective. *Annu. Rev.*
654 *Phytopathol. Vol 53* **53**, 513–539 (2015).
- 655 63. Villari, C., Herms, D. A., Whitehill, J. G. A., Cipollini, D. & Bonello, P. Progress and gaps in

understanding mechanisms of ash tree resistance to emerald ash borer, a model for wood-boring insects that kill angiosperms. *New Phytol.* **209**, 63–79 (2016).

64. Magoč, T. & Salzberg, S. L. FLASH: Fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**, 2957–2963 (2011).

65. Leggett, R. M., Clavijo, B. J., Clissold, L., Clark, M. D. & Caccamo, M. Next clip: An analysis and read preparation tool for nextera long mate pair libraries. *Bioinformatics* **30**, 566–568 (2014).

66. Clavijo, B. J. *et al.* W2RAP: a pipeline for high quality, robust assemblies of large complex genomes from short read data. *bioRxiv* (2017). doi:10.1101/110999

67. Luo, R. *et al.* SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* **1**, 18 (2012).

68. Tarailo-Graovac, M. & Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinforma.* 1–14 (2009). doi:10.1002/0471250953.bi0410s25

69. Wu, T. D. & Watanabe, C. K. GMAP: A genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* **21**, 1859–1875 (2005).

70. Stanke, M. & Morgenstern, B. AUGUSTUS: A web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res.* **33**, 465–467 (2005).

71. Emanuelsson, O., Brunak, S., von Heijne, G. & Nielsen, H. Locating proteins in the cell using TargetP, SignalP and related tools. *Nat. Protoc.* **2**, 953–971 (2007).

72. Enright, A. J., Van Dongen, S. & Ouzounis, C. A. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* **30**, 1575–1584 (2002).

73. Jorda, J. & Kajava, A. V. T-REKS: Identification of Tandem REpeats in sequences with a K-meanS based algorithm. *Bioinformatics* **25**, 2632–2638 (2009).

74. Ceroni, A., Passerini, A., Vullo, A. & Frasconi, P. Disulfind: A disulfide bonding state and cysteine connectivity prediction server. *Nucleic Acids Res.* **34**, 177–181 (2006).

75. Choi, J. *et al.* Fungal secretome database: integrated platform for annotation of fungal secretomes. *BMC Genomics* **11**, 105 (2010).

- 682 76. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*
683 **1303.3997v**, 1–3 (2013).
- 684 77. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and
685 population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993
686 (2011).
- 687 78. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
- 688 79. Delaneau, O., Zagury, J.-F. & Marchini, J. Improved whole-chromosome phasing for disease and
689 population genetic studies. *Nat. Methods* **10**, 5–6 (2013).
- 690 80. Kasuga, T., White, T. J. & Taylor, J. W. Estimation of nucleotide substitution rates in Eurotiomycete
691 fungi. *Mol. Biol. Evol.* **19**, 2318–2324 (2002).
- 692 81. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features.
693 *Bioinformatics* **26**, 841–842 (2010).
- 694 82. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide
695 polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w 1118; iso-2; iso-3.
696 *Fly (Austin)*. **6**, 80–92 (2012).
- 697 83. Librado, P. & Rozas, J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism
698 data. *Bioinformatics* **25**, 1451–1452 (2009).
- 699 84. McKenna, A. *et al.* The Genome Analysis Toolkit: A MapReduce framework for analyzing next-
700 generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
- 701 85. Yang, Z. H. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591
702 (2007).
- 703 86. Bowman, A. W. & Azzalini, A. R package ‘sm’: nonparametric smoothing methods (version 2.2-5.4).
704 (2014).
- 705 87. Huson, D. H. & Bryant, D. Application of Phylogenetic Networks in Evolutionary Studies. *Mol. Biol.*
706 *Evol.* **23**, 254–267 (2006).
- 707 88. Krzywinski, M. *et al.* Circos: an Information Aesthetic for Comparative Genomics. *Genome Res* **19**,

708 1639–1645 (2009).

709 89. Wey, T., Schlegel, M., Stroheker, S. & Gross, A. MAT- gene structure and mating behavior of

710 Hymenoscyphus fraxineus and Hymenoscyphus albidus. *Fungal Genet. Biol.* **87**, 54–63 (2016).

711 90. Semagn, K., Babu, R., Hearne, S. & Olsen, M. Single nucleotide polymorphism genotyping using

712 Kompetitive Allele Specific PCR (KASP): Overview of the technology and its application in crop

713 improvement. *Mol. Breed.* **33**, 1–14 (2014).

714 91. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic Local Alignment Search

715 Tool. *J. Mol. Biol.* **215**, 403–410 (1990).

716 92. Excoffier, L. & Foll, M. fastsimcoal: A continuous-time coalescent simulator of genomic diversity

717 under arbitrarily complex evolutionary scenarios. *Bioinformatics* **27**, 1332–1334 (2011).

718

719