FEMS Microbiology Ecology, 94, 2018, fix168

doi: 10.1093/femsec/fix168
Advance Access Publication Date: 6 December 2017
Research Article

RESEARCH ARTICLE

Co-invading symbiotic mutualists of *Medicago polymorpha* retain high ancestral diversity and contain diverse accessory genomes

Stephanie S. Porter[1,*,†], Joshua J. Faber-Hammond[1] and Maren L. Friesen[2,3,4]

[1]School of Biological Sciences, Washington State University, 14204 NE Salmon Creek Ave, Vancouver, WA 98686, USA, [2]Department of Plant Biology, Michigan State University, 612 Wilson Road, East Lansing, MI, 48824, USA, [3]Department of Plant Pathology, Washington State University, P.O. Box 646430 Pullman, WA 99164, USA and [4]Department of Crop and Soil Sciences, Washington State University, P.O. Box 646420 Pullman, WA 99164, USA

*Corresponding author: Washington State University, School of Biological Sciences, 14204 NE Salmon Creek Ave, Vancouver, WA 98686, USA. Tel: (360) 546-9478; E-mail: stephanie.porter@wsu.edu

One sentence summary: The beneficial nitrogen-fixing symbionts that facilitate a plant's biological invasion appear to have spread from Europe with their host, retain high levels of genomic diversity and harbor novel genes.

Editor: Angela Sessitsch

†Stephanie S. Porter, http://orcid.org/0000-0002-1193-9705

ABSTRACT

Exotic, invasive plants and animals can wreak havoc on ecosystems by displacing natives and altering environmental conditions. However, much less is known about the identities or evolutionary dynamics of the symbiotic microbes that accompany invasive species. Most leguminous plants rely upon symbiotic rhizobium bacteria to fix nitrogen and are incapable of colonizing areas devoid of compatible rhizobia. We compare the genomes of symbiotic rhizobia in a portion of the legume's invaded range with those of the rhizobium symbionts from across the legume's native range. We show that in an area of California the legume *Medicago polymorpha* has invaded, its *Ensifer medicae* symbionts: (i) exhibit genome-wide patterns of relatedness that together with historical evidence support host-symbiont co-invasion from Europe into California, (ii) exhibit population genomic patterns consistent with the introduction of the majority of deep diversity from the native range, rather than a genetic bottleneck during colonization of California and (iii) harbor a large set of accessory genes uniquely enriched in binding functions, which could play a role in habitat invasion. Examining microbial symbiont genome dynamics during biological invasions is critical for assessing host-symbiont co-invasions whereby microbial symbiont range expansion underlies plant and animal invasions.

Keywords: co-invasion; microbial invasion; invasion genomics; rhizobia; legume; mutualism

INTRODUCTION

Biological invasions are a major driver of species diversity loss, leading to extinctions and substantial economic losses (Vitousek, D'Antonio and Loope 1997; Mack *et al.* 2000; Pimentel 2011; Lowry *et al.* 2013). Invasive plants and animals associate with diverse microbiota as they spread (van der Putten, Klironomos and Wardle 2007; Traveset and Richardson 2014). While pathogens amplified by invasive species have been well documented (Vitousek, D'Antonio and Loope 1997; Mack *et al.*

Received: 14 September 2017; **Accepted:** 29 November 2017

2000; Pimentel *et al.* 2001; Parker and Gilbert 2004; Inderjit and van der Putten 2010; Pimentel 2011; Strayer 2012; Lowry *et al.* 2013), mutualistic symbionts amplified by invasive hosts are relatively understudied (Desprez-Loustau *et al.* 2007; van der Putten, Klironomos and Wardle 2007; Pringle *et al.* 2009; Litchman 2010; Traveset and Richardson 2014), even though these mutualists can define host niches (Marquez *et al.* 2007; Oliver *et al.* 2010; Friesen *et al.* 2011; Nougué *et al.* 2015) and host invasion biology (Pringle *et al.* 2009; Vilcinskas *et al.* 2013; Traveset and Richardson 2014). For hosts that associate with cosmopolitan microbiota (Finlay 2002; De Wit and Bouvier 2006), symbiont availability does not constrain invasions. However, dispersal limitation and environmental filtering can limit microbial distributions (Hanson *et al.* 2012), which can restrict host ranges (Parker 2001; Nuñez, Horton and Simberloff 2009; Stanton-Geddes and Anderson 2011; Hayward *et al.* 2015; Hayward, Horton and Nuñez 2015; Simonsen *et al.* 2017). While symbionts are easily overlooked as 'invisible' components of invasion (Amsellem *et al.* 2016), high throughput microbial genomics offers an opportunity to reveal symbiont invasion biology.

The invasion genetics of mutualistic symbionts are relatively understudied (Litchman 2010; though see Brown *et al.* 2014 and Harrison *et al.* 2017), but can give critical insights into the invasion process. Invaded range populations may have lower genetic diversity than native ones if colonization resulted from limited, long-distance dispersal (Dlugosch and Parker 2008; Lachmuth, Durka and Schurr 2010; Zhang, Zhang and Barrett 2010). Alternatively, invaded range genetic diversity may equal or exceed that in the native range due to multiple invasions or post-colonization admixture (Dlugosch and Parker 2008; Lachmuth, Durka and Schurr 2010; Puzey and Vallejo-Marín 2014; Barrett 2015). Because horizontal transfer events can result in contrasting evolutionary dynamics for different portions of microbial genomes (Bailly *et al.* 2011; Polz, Alm and Hanage 2013), genome-wide studies offer a more powerful approach than studies that investigate only a handful of genes (Silva *et al.* 2005; Crisóstomo, Rodríguez-Echeverría and Freitas 2013; Horn *et al.* 2014; Nguyen, Spooner-Hart and Riegler 2016). Furthermore, horizontally transmitted accessory genes comprise a reservoir of adaptive variation that may propel microbial adaptation to novel environments (Polz, Alm and Hanage 2013; Porter *et al.* 2016; Young 2016), which can be a critical facet of invasion biology.

Population genetic comparisons can distinguish among four invasion scenarios. First, familiar symbionts from the host's native range could co-invade a new area with their host, as has occurred for a number of mycorrhizae (Hayashi *et al.* 2001; Dickie *et al.* 2010; Nuñez and Dickie 2014; Bogar, Dickie and Kennedy 2015; Hayward *et al.* 2015). Here, the same symbiont genotypes are present both ranges, but historical records indicate that symbionts were absent in the invaded range prior to host introduction (Pringle *et al.* 2009). Second, major genomic rearrangements (chimerism) and horizontal gene transfer (HGT) of genomic islands can give rise to evolutionary innovations that help microbes colonize novel environments (Ochman, Lawrence and Groisman 2000; Kuo and Ochman 2009; Litchman 2010; Darmon and Leach 2014). For example, European *Lotus corniculatus* colonized New Zealand in concert with the transfer of the symbiosis island from its European rhizobial partner, *Mesorhizobium loti*, into a non-nodulating native New Zealand *Mesorhizobium* (Sullivan *et al.* 1995, 1996). Third, hosts can switch partners to acquire novel symbionts in the invaded range. In this case, hosts are symbiotic generalists that associate with different symbiont genotypes in the native and invaded range, as occurs for *Cytisus scoparius* and *Mimosa pigra* and their nitrogen-fixing symbionts (Lafay and Burdon 2006; Parker, Wurtz and Paynter 2006). Finally, if symbionts are cosmopolitan, the same symbiont genotypes are available to the host in novel and native ranges, as occurs in plants that associate with cosmopolitan arbuscular mycorrhizal fungi (Moora *et al.* 2011). Here, there would be no bottleneck as expected in co-invasions. In some cases, symbionts could invade first, spread on native alternative hosts (Wolfe *et al.* 2010; Bonito *et al.* 2011), and then facilitate the spread of invasive hosts. Understanding the prevalence of these mechanisms of symbiont invasion is critical for integrating microbes into invasion biology.

Leguminous plants comprise some of the most noxious invasive species (Daehler 1998; Richardson *et al.* 2000), but we are only beginning to understand the role of their rhizobium symbionts in invasion (Traveset and Richardson 2014; Simonsen *et al.* 2017; LaPierre *et al.* 2017). Rhizobia are horizontally transmitted bacteria that persist between hosts as soil saprobes (Narożna *et al.* 2015) and rhizobial limitation can reduce the spread of invasive plants (Parker 2001; Parker, Malek and Parker 2006; Simonsen *et al.* 2017). Plants house rhizobia intracellularly in root nodules where they provide reduced nitrogen in exchange for photosynthates (Oldroyd 2013); the genes underlying these processes cluster on transmissible plasmids or islands (Galibert *et al.* 2001; Reeve, Chain and O'Hara 2010; Sugawara *et al.* 2013; Haskett *et al.* 2016). *Medicago polymorpha* originates from Mediterranean Europe and North Africa and is a globally distributed invader (Lesins and Lesins 1979; Silva, Kan and Martinez-Romero 2007). It is specialized on the rhizobium *Ensifer medicae,* and also rarely nodulates with *Ensifer meliloti* in its native range (Rome *et al.* 1996). This study provides a genome-wide perspective on rhizobia nodulating *M. polymorpha* at one reserve in its invaded range. We compare rhizobium genomes from this site to published genomes from the same species of rhizobium across the native range of its host to ask: (i) Do genome-wide patterns of relatedness suggest *M. polymorpha* symbionts in the sampled portion of the invaded range are novel chimeras, co-invading species, novel partner species or cosmopolitan generalist symbionts? (ii) Do population genomics support reductions in symbiont diversity in the invaded range, as would be predicted by a colonization bottleneck? And (iii) Do invaded range symbiont genomes harbor unique accessory genes that could have been acquired during adaptation to their new habitat?

METHODS

Sample collection

Seventeen *Ensifer medicae* strains were isolated from one root nodule from each of 17 field-collected *Medicago polymorpha* plants from the McLaughlin Natural Reserve in California, USA (38.8709975, -122.419128) (Table S1, Supporting Information), as described in Porter and Rice (2013).

Genomic sequencing

Californian isolates were grown in 10 mL tryptone yeast broth cultures at 30°C shaking at 200 rpm until dense, and then cells pelleted and frozen at −80°C. DNA extraction and genomic library construction followed Dunham & Friesen (2013) with modifications as in Porter *et al.* (2016). Samples were quantified using a Qubit fluorometer (Thermo Fisher, Waltham, MA) and ∼500 ng of DNA was used for genomic library construction. Libraries were sequenced in paired 76 bp format on an Illumina GAIIx at the University of Southern California.

In total, 17 *E. medicae* genome assemblies from across the native range of *M. polymorpha* were downloaded from NCBI and MicroScope assembly databases (Epstein *et al.* 2012; Sugawara *et al.* 2013; Reeve *et al.* 2015). These strains originated from France, Italy, Greece, Jordan, Iraq and Syria, which span the Mediterranean region of Europe and the Middle East where *M. polymorpha* is native (Lesins and Lesins 1979). *Ensifer meliloti* strain Rm1021 (Galibert *et al.* 2001), the sister species of *E. medicae* (Bailly *et al.* 2006, 2011; Sugawara *et al.* 2013), was used as an outgroup for phylogenetic analyses.

Assembly and annotation

Californian draft genomes were assembled using the A5 pipeline (Tritt *et al.* 2012), using the May 18th 2012 Linux x64 A5 version and run with default parameters. Assembly statistics, including genome size, number of contigs and average contig length were calculated using Prinseq v0.20.4 (Schmieder and Edwards 2011; Table S1, Supporting Information). All 35 genome assemblies (including *E. meliloti*) were submitted to RAST v2.0 for functional annotation by RASTtk (Aziz *et al.* 2008).

To obtain gene ontology (GO) terms for consensus sequences for each orthogroup, we created consensus sequences for each orthogroup using the EMBOSS v6.5.7 cons tool running on protein sequences with default parameters (Rice, Longden and Bleasby 2000). Consensus amino acids were called at a locus if it was present in at least 50% of orthogroup sequences. Consensus sequences were mapped and annotated with BLAST2GO v3.2 using the top five BLASTP hits from NCBI's nr database (Ana Conesa 2008) and BLAST2GO's embedded InterProScan search. This allowed us to extract and analyze top BLAST hits and GO information for 6858 out of the 12 138 genes in the pangenome.

Core and accessory genomic variation

Orthologs were identified using OrthoMCL (Wang *et al.* 2003) with the inflation value for MCL clustering at 1.5 and other parameters at default values. Copy numbers of each orthogroup by strain indicate the number of core vs. accessory and single copy vs. duplicate genes. Core genes were defined to have at least one ortholog copy in all *E. medicae* strains examined, while genes with variable presence/absence rates were considered accessory genes, excluding singletons. Singletons were defined as genes found in only one *E. medicae* strain. Additionally, we used the following definitions in enrichment tests: (i) *private core genes* are present in all strains in one geographic range, but accessory in the other range, while (ii) *private accessory genes* were accessory in one range and completely absent in the other. There were no orthogroups that were core in one range and completely absent the other.

To identify SNPs in the core genome, ClustalOmega v1.1.0 was used to align orthogroup amino acid sequences from all single-copy core genes using default parameters. Orthogroups with more than one copy in a strain were excluded from this analysis to maintain a balance in copy number for downstream statistical tests. Nucleotides were substituted for their respective translated protein sequences from each multiple sequence alignment using pal2nal.pl (http://www.bork.embl.de/pal2nal/). Nucleotide MSAs were sorted by replicon, and SNPs were then called for each set using parseSNPs.py (https://gist.github.com/peterk87/8409706). The SNP fasta output file was reformatted through custom scripts and PGDSpider v2.0.9.1 for use in STRUCTURE v2.3.0, Genepop v4.1.2 and SplitsTree v4.14.2 (Raymond and Rousset 1995; Huson 1998; Pritchard, Stephens and Donnelly 2000; Hubisz *et al.* 2009; Lischer and Excoffier 2012).

To estimate the genomic organization of core and accessory variation in *E. medicae*, we used the finished WSM419 genome that consists of a chromosome (3.78 Mbp), two mega-plasmids: psmed01 (1.57 Mbp), psmed02 (1.25 Mbp) and an additional plasmid, psmed03 (0.22 Mbp) (Reeve, Chain and O'Hara 2010). Representative gene sequences from each orthogroup were mapped to WSM419 using default parameters in Bowtie2 v2.2.6 (Langmead and Salzberg 2012) and filtered to include only those with primarily unique loci in the genome (Bowtie2 mapping scores $\geq$ 30). Unmapped sequences are reported as unknown loci.

To determine whether Californian *E. medicae* have lost major portions of the genome relative to native range strains, we visualized the genome content using BlastRing (BRIG) v0.95 (Alikhan *et al.* 2011). The WSM419 *E. medicae* reference genome was divided into individual replicons and the 34 genome assemblies were BLASTed to this reference. Rings were organized by native versus invaded range in California for visualization.

Patterns of relatedness

To determine evolutionary relationships between *E. medicae* strains from the invaded and native host range, we built restricted maximum likelihood trees and neighbor nets for (i) SNPs in single-copy core genes in each replicon, and (ii) accessory gene content. Analyses used all 17 native and 17 invaded range strains and the sister species, *E. meliloti,* as the outgroup. Trees were constructed in RAxML v8.0.6 (Stamatakis 2014) using concatenated single copy core gene sequences. For the chromosome, we concatenated three sets of 500 randomly selected genes of the 1831 total genes to check for possible tree disparities. For psmed01 and psmed02, we concatenated all 603 and 89 genes, respectively, and for psmed03, no genes were core. Nucleotide alignments converted from protein MSAs with pal2nal.pl (as described above) were concatenated using FASconCAT v1.0 (Kück and Meusemann 2010). Accessory gene input data were in binary presence/absence format for 7878 orthogroups. The RAxML model used for nucleotide sequences was GTRGAMMA and for accessory genes BINGAMMA, both with 1000 iterations and 100 bootstraps. We also built neighbor nets using SplitsTree v4.14.2 (Huson 1998) based on (i) SNPs segregating in the core genome, and (ii) accessory gene presence/absence.

Population genetic diversity and population structure

To assess whether there is evidence of a population bottleneck during invasion, we calculated expected heterozygosity (*He*) in each range. This analysis focused on all non-singleton SNPs in the core genome and used Genepop Software v4.1.2 (Raymond and Rousset 1995). Heterozygosity, *He*, was calculated for each polymorphic site using the allele frequency for each geographic range separately and for the total population. Changes in population-level *He* rates in California compared to the native range were compared via two-tailed *T*-tests using a Bonferroni corrected *P*-value of 0.0125 for four tests (each replicon and across genome). Private SNP alleles were tallied for each replicon.

Fst was calculated at each site in the core genome using the following formula: $Fst = (Ht - Hs)/Ht$. *Ht* represents *He* calculated for the total sampled population of 34 *Ensifer* strains and *Hs* is the average of individual *He* values from each of the two ranges. *Fst* values were averaged for each replicon and across the genome.

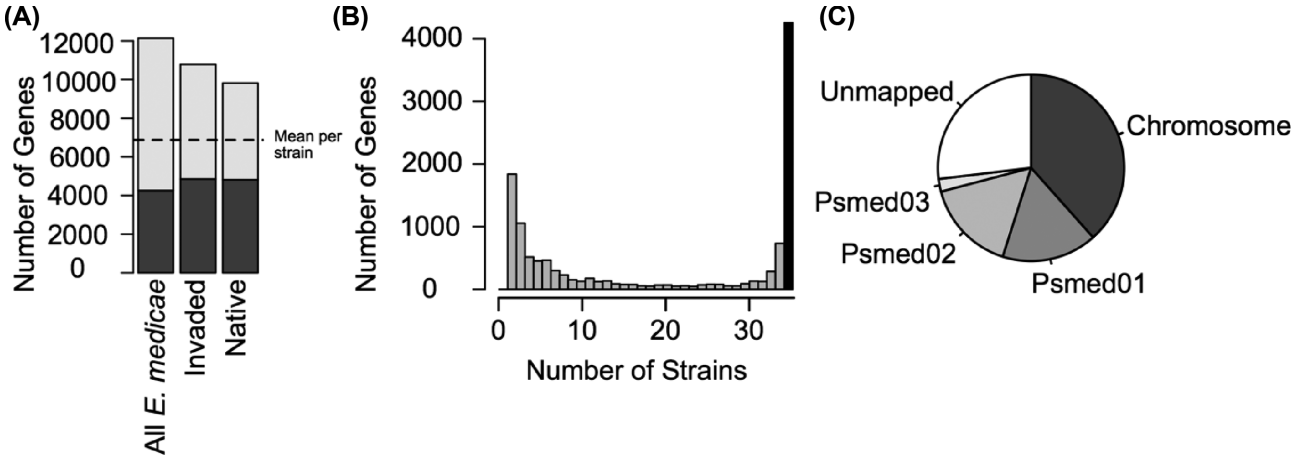

Figure 1. The majority of annotated genes map to the *E. medicae* WSM419 reference genome. (**A**) Pangenome size for all 34 *E. medicae* strains, 17 strains from the sampled portion of the invaded range and 17 strains from across the native range. Core genes, dark gray; accessory genes, light gray. (**B**) Distribution of core (black) and accessory (gray) genes across strains. (**C**) Of the 12 268 orthogroups in the pangenome, 72.2% aligned to sequences in the chromosome or Psmed01–Psmed03 replicon in the WSM419 reference genome.

In order to define patterns of genetic diversity in the invaded range site, we ran STRUCTURE on the core genome SNPs and on accessory genome presence/absence on the 17 Californian strains, with the burnin period and MCMC repetitions set at 10 000 and 15 iterations, respectively, and *K* ranging from 1 to 12. We determined the best *K* by identifying peak values of ΔK when plotted against *K* (Evanno, Regnaut and Goudet 2005).

Unique gene content and functional implications

To determine whether Californian or native range *E. medicae* harbor unique genes, we examined private core and accessory pangenomic compartments by range. We also identified genes with variable frequencies between ranges by conducting two-tailed *T*-tests on the copy number for each orthogroup. We calculated the proportion of orthogroups with duplicates in at least one strain, number of accessory vs. core genes per genome for range comparisons and ran functional enrichment tests on pangenome subsets.

To examine the functional implications of private accessory genomic compartments, we conducted enrichment tests. We primarily conducted two analyses using the private accessory orthogroups from each range as the test sets and the remaining orthogroups in the pangenome as the reference set for each comparison. Singletons were excluded from this analysis since they may be partial sequences resulting from incomplete genome assembly and thus skew functional annotation results. Enrichment analysis on GO terms was implemented via Fisher's Exact Test through BLAST2GO with a False Discovery Rate of 0.05. We also tallied top species BLAST hits for private accessory genes in each range to address whether range-exclusive genes have their origin as rare *Ensifer* genes or from HGT.

RESULTS

Assembly and annotation statistics

Draft genomes from the McLaughlin Reserve in the invaded range (excluding Str5) assembled into an average of 276 contigs with a mean size of 31.7 kb for a total mean genome size of 6.8 Mb (Table S1, Supporting Information). In contrast, native range genomes (excluding completed reference genome, WSM419) assembled into an average of 196 contigs with an average length of 48.5 kb for a total mean genome size of 6.9 Mb. Despite these similar metrics, it is possible that differences in sequencing depth (~20X vs ~108X coverage (Epstein *et al.* 2012)) resulted in rare sequencing errors being retained more often in invaded range strains. The Str5 assembly resulted in an unusually large number of contigs (2373), low average contig size (2.4 kb), and slightly smaller genome size (5.8 Mb), yet has a comparable number of RAST annotated features with other strains. Annotated genomes (all strains in both ranges) have an average of 7285 transcribed features distributed amongst 6878 orthogroups/genes. The average genome contains 6574 single-copy genes (Fig. 1A) and 304 genes that were duplicated at least once in the genome based on OrthoMCL assignments (Fig. S1, Supporting Information).

There are more single-copy orthogroups per strain in the invaded range than in the native range (a mean of 6845 vs. 6302; $T_{16} = 6.22$, $P = 1.22E{-}05$), while the number of multi-copy orthogroups did not differ between ranges (a mean of 291 vs. 316; $T_{16} = -0.89$, $P = 0.39$) (Fig. 1A). In total, there are 12 138 unique orthogroups in pangenome of the selected strains of *E. medicae* (Table S2, Supporting Information). The mode value of gene copies per orthogroup is 34 (for 2905 orthogroups), primarily representing single copy genes found in all 34 strains (Fig. 1B). Excluding singletons, there is an abundance of rare accessory orthogroups (3780) found in five or fewer strains. Each strains harbors an average of 269 singleton (i.e. unique) orthogroups, with Str5 yielding the highest number of singletons (1534), many of which are partial gene fragments likely derived from assembly artifacts. Without singletons, Str5 showed a comparable number of genes with closely related strains; therefore, singletons from all strains were excluded from downstream analyses to retain only conservative gene calls.

Of the 12 138 orthogroups in the *E. medicae* pangenome, 73% of sequences mapped to a replicon in the WSM419 reference genome (Fig. 1C), but a portion of these yielded low mapping scores (<30) indicating they either had weak alignments or were repetitive sequences that mapped to multiple locations. A majority of sequences (67.3%) aligned to the reference genome with strong Bowtie2 mapping scores (≥30), while 4098 genes (33.8%) were not confidently mapped to any molecule. A majority of private accessory genes (2752/3804; 72.3%) from both ranges remained unmapped. All 4260 orthogroups core to all 34 strains

were found in the mapped portion of the pangenome since they include an ortholog in WSM419.

Patterns of relatedness

Ensifer medicae in the sampled portion of the invaded range in California are primarily a derived subset of native range diversity based on patterns of relatedness at SNPs in core genes. Californian *E. medicae* appears to be comprised of three major clades (Clade 1, 2 and 3) based on phylogenies estimated with RAxML (Fig. 2), population groupings from STRUCTURE (Fig. S2, Supporting Information), neighbor nets estimated in SplitsTree (Figs S3 and S5, Supporting Information) and clustering of strains by accessory gene presence/absence data (Fig. S6, Supporting Information). Among the 32 genomes in our study with > 4 contigs, contig size (a measure of assembly quality) is weakly negatively correlated ($R = -0.53$, $T_{30} = -3.5$, $P < 0.01$) with branch length (Fig. S7, Supporting Information). Str5 has a low number of recovered reads, and numerous small contigs, which can drive the overestimation of sequence divergence from other strains and likely underlie its long branch lengths. Due to the random nature of sequencing errors, these are unlikely to affect tree or network topology, which we confirmed by building neighbor networks and conducting STRUCTURE analysis both with and without Str5.

Within a clade, Californian strains tend to exhibit consistent patterns of close relatedness across replicons, with the exception of four strains (Str7, Str8, Str9 and Str18) that group in different clades for different replicons (Fig. 2). The relationship among clades, however, is different on different replicons (Fig. 2). The psmed01 phylogenetic tree and neighbor net both show Clade 1 as basal and closest to the *E. meliloti* outgroup (Fig. S4, Supporting Information), which may point to a unique origin of the Clade 1 psmed01 plasmid. Neighbor nets with and without Str5 yielded similar patterns of relatedness among the other 33 strains, confirming that the inclusion of this low-coverage assembly has little impact on overall strain clustering (Figs S3 and S8, Supporting Information).

STRUCTURE indicates there are either two or three subpopulation clusters in the invaded range depending on the molecule examined (Fig. S2, Supporting Information). Among Californian strains, peaks of ΔK indicate four subpopulation clusters ($K = 4$) for the chromosome and accessory gene presence/absence, and two ($K = 2$) for psmed01 and psmed02. However, for the chromosome and accessory gene presence/absence, the four optimum clusters consist of the three expected clusters based on phylogenetic data, and the outlier Str5. STRUCTURE analysis repeated with Str5 removed best supports three subpopulations ($K = 3$) for the chromosome and accessory data, which correspond to the clades observed in the RAxML phylogenies and SplitsTree networks.

Major genomic loss events

Californian *E. medicae* do not appear to have lost major portions of the genome relative to native range strains. Visual comparisons of invaded and native range genomes using BlastRing alignments to the WSM419 genome revealed that all genomes from both ranges contain the majority of genes contained within the WSM419 chromosome, psmed01 and psmed02 replicons (Fig. 3). However, there are substantial differences in the genomic architecture of psmed03 between all strains and the WSM419 reference, which contains a large region of tandem duplications. Invaded range genomes contain 89% of the annotated genes in the WSM419 reference genome, while native range genomes contain an average of 88.4%. In comparison, psmed03 was highly variable with an average of 63.7% of the 157 WSM419 genes found in invaded range strains and 43.2% found in native range strains.

Population genetic diversity and population structure

Our results support a scenario in which *E. medicae* did not undergo a major population bottleneck during co-invasion with their plant host. The average expected *He* from each locus differed between ranges (Bonferroni correction for four tests; $P < 0.0125$; Table 1). *He* is greater in the invaded than in the native range for loci located on the WSM419 chromosome (+30.9%; $T_{6,280} = 21.18$, $P = 3.05E{-}96$), WSM419 psmed01 (+140.8%; $T_{6,396} = 62.38$, $P < 0.0001$), and the genome overall (+71.6%; $T_{13,\ 339} = 55.99$, $P < 0.0001$), although not for psmed02. The *Fst* between ranges is 0.0983 (Table 1). Psmed01 is the most divergent replicon ($Fst = 0.1301$) and the chromosome is least divergent with $Fst = 0.0671$. Invaded range genomes harbor more private alleles overall (5511), than do native range strains (2445). Compared to native range strains, invaded range strains have 100% and 211% more private alleles on the chromosome and psmed01 replicon, respectively, but 30% fewer private alleles on the psmed02 replicon. Divergence between native and invaded range strains appears to be largely driven by private alleles on psmed01 from Clade 1. *He* on this plasmid is twice that of native range strains, and Clade 1 harbors over half of invaded range private alleles.

To exclude SNPs resulting from sequencing errors, the metrics above are calculated after excluding singleton variable sites. This does not appear to overestimate *He* and *Fst* because differences in *He* and private allele count are even greater when singleton SNPs, many of which are likely real mutations and rare alleles, are included (Table S3, Supporting Information).

Unique gene content and functional implications

A large set of core genes is shared between California and the native range. 4260 (35%) orthogroups are core, while 7878 (65%) are accessory. California strains contained more accessory genes (mean of 2877) than did native range strains (mean of 2359) ($T_{16} = 5.51$, $P = 4.75E{-}05$). In particular, all strains in Californian Clade 1 contain more accessory genes than any other strains (mean of 3276) (Fig. S9, Supporting Information).

Invaded range strains in California harbor more private accessory genes than do native range strains (Fig. 4). There are no private core genes in either range that were completely absent from the other range, suggesting that the core genomes are relatively stable. Ignoring the four mosaic genomes (Str7, Str8, Str9 and Str18), 857 of the 2319 total private accessory genes in California were exclusive to Clade 1 (six strains), while Clades 2 (three strains) and 3 (four strains) had 276 and 348 exclusive orthogroups, respectively. Californian strains have a larger number of orthogroups with significantly higher gene frequencies than do invaded range strains (Table 2). After Bonferroni correction, 18 orthogroups show higher frequency in the invaded range but are absent in most native samples, and 10 orthogroups show higher frequency in the native range.

Private accessory genes in the invaded range are enriched for a variety of 'binding' molecular functions (Table S4 and Fig. S10, Supporting Information). Private accessory genes in the native range are enriched for transposon-related genes, other repetitive element-associated GOs and some binding-related GOs. Private accessory genes show a majority of top BLAST hits to

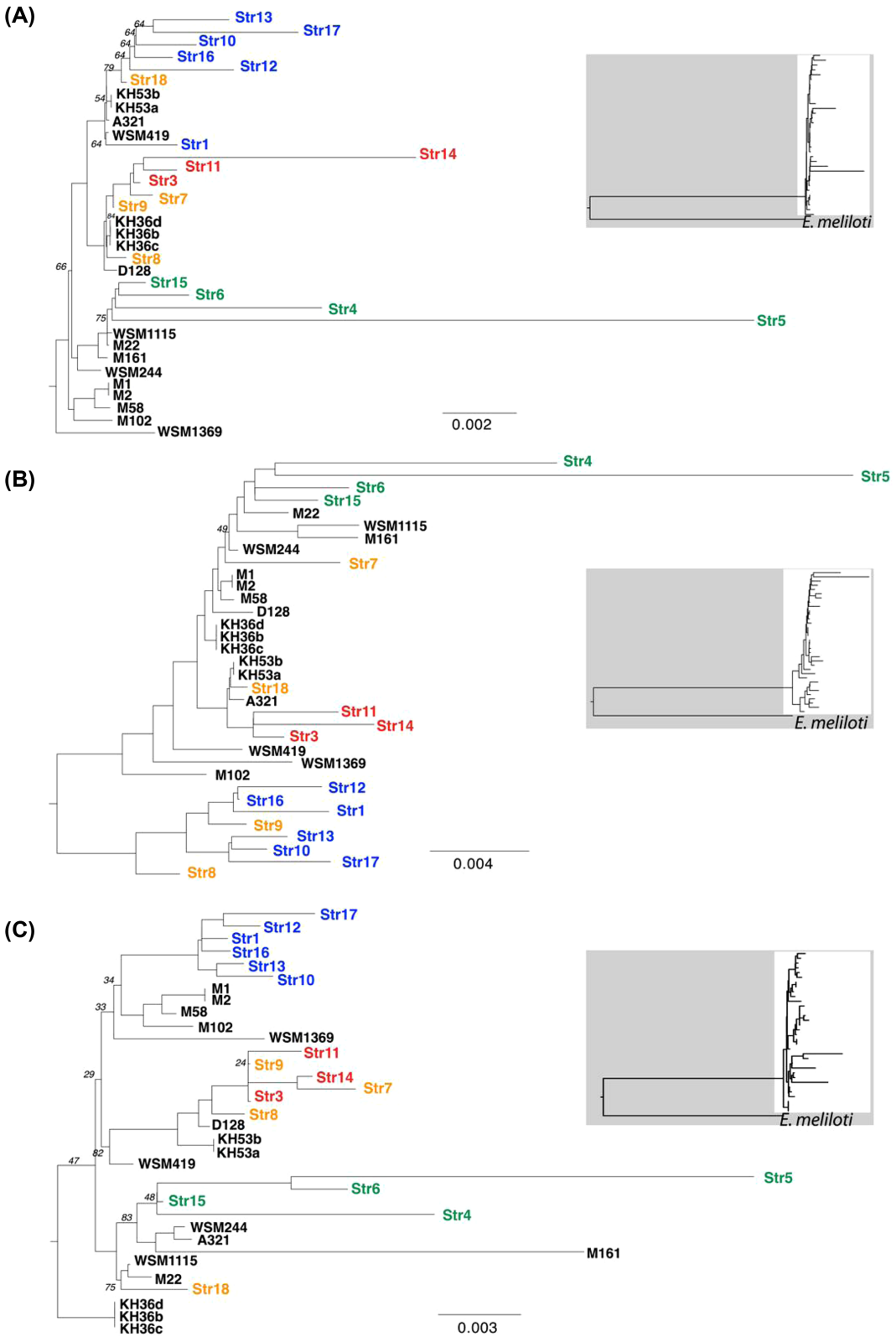

Figure 2. Major clades of diversity present across replicons in the native range of *E. medicae* have representatives in the invaded range in California. Phylogeny estimated for *E. medicae* strains using restricted maximum likelihood in RAxML, based 500 randomly selected single copy core genes from: (**A**) the chromosome, (**B**) the psmed01 replicon and (**C**) the psmed02 replicon. Strains from the invaded range in California are color coded by cluster as determined by STRUCTURE; blue, Clade 1; red, Clade 2; green, Clade 3; yellow, mosaic genomes. Relationship to the sister species, *E. meliloti*, is shown in inset figures. All nodes had bootstrap values >85 except for those shown in the figure.

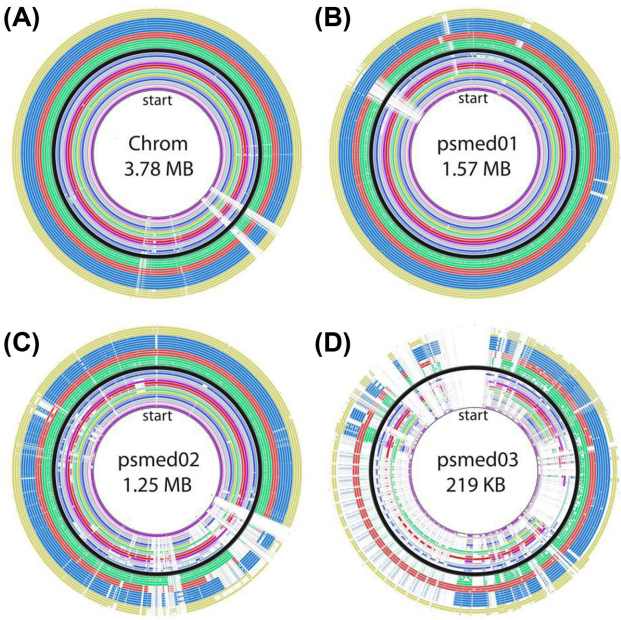

Figure 3. Californian *E. medicae* do not appear to have lost major portions of the genome relative to native range strains. BlastRing figures show coverage of focal *E. medicae* strain genomes across the **(A)** chromosome, **(B)** psmed01, **(C)** psmed02 and **(D)** psmed03 of the WSM419 reference *E. medicae* genome (black ring) for 17 strains from host *Medicago polymorpha*'s native range (rings inside black ring) and 17 strains from *M. polymorpha*'s invaded range in California (rings outside of black ring). Invaded range strains are colored according to clustering patterns in phylogenetic trees and STRUCTURE analysis; blue, Clade 1; red, Clade 2; green, Clade 3 and yellow, mosaic genomes.

Ensifer species (Fig. S11, Supporting Information), which could indicate many of these genes are rare orthologs or paralogs of genes that have been sequenced in strains of *E. medicae* or *E. meliloti*. A small portion of genes have stronger homology with genes from non-*Ensifer* species and may have been acquired through HGT. Californian private accessory genes exclusive to invaded Clade 1 are enriched for DNA recombination, transposase activity, DNA mediated transposition, transposition and DNA metabolic process (Table S4, Supporting Information). Invaded range Clade 2 and 3 private accessory genes did not yield enriched GO terms.

DISCUSSION

Beneficial symbionts play a critical role in the invasion biology of exotic plants and animals (Richardson *et al.* 2000; Pringle *et al.* 2009; Himler *et al.* 2011; Traveset and Richardson 2014). However, the invasion genomics of most beneficial microbes remains a black box (van der Putten, Klironomos and Wardle 2007; Litchman 2010; Le Roux *et al.* 2017). Using population genomics, we show that in an area the legume *M. polymorpha* has invaded, its *E. medicae* symbionts: (i) likely co-invaded with their host from Europe into California, (ii) did not experience a genetic bottleneck during colonization of California, but rather show the introduction of the majority of native range genetic diversity and (iii) harbor a large set of accessory genes uniquely enriched in binding functions, which could play an adaptive role in colonization.

Symbiotic invasion scenarios

Our data support a scenario of co-invasion of *E. medicae* symbionts with their host, *M. polymorpha*. Approximately 89% of the gene content from the *E. medicae* WSM419 reference genome (Galibert *et al.* 2001; Sugawara *et al.* 2013) is present in symbiont genomes in both the native and invaded host range. This high similarity supports the conclusion, based on *16S* or MLST sequences, that *M. polymorpha* associates with the native range *E. medicae* species in their invaded ranges in California (Silva, Kan and Martinez-Romero 2007; Porter and Rice 2013). In addition, Californian strains from the three major clades are largely derived from *E. medicae* clades present across *M. polymorpha*'s native range. This pattern varies across the genome: patterns of relatedness among SNPs in the chromosome and psmed02 replicon support a scenario of a dispersal of *E. medicae* from *M. polymorpha's* native range to California, while those in the psmed01 replicon suggest Californian strains from Clade 1 could have a unique origin or derive from an unsampled native range *E. medicae* strain. In psmed03, a highly variable small plasmid (213 kb in WSM419), patterns of gene presence and absence in Californian strains suggest that this replicon is derived relative to native range strains, though psmed03 contains only accessory genes, precluding an SNP-based analysis of relatedness.

Our data do not support a scenario of genomic chimerism facilitating *E. medicae* invasion, which would require symbiosis genes transferring into a novel chromosomal background. This is in contrast to cases where genomic chimerism events within *Mesorhizobium* and *Bradyrhizobium* have facilitated range expansion of *L. corniculatus* (Sullivan *et al.* 1995; Sullivan and Ronson 1998), *Biserrula pelecinus* (Nandasena *et al.* 2007) and *C. scoparius* (Horn *et al.* 2014). In the *M. polymorpha* invasion, Californian *E. medicae* are derived from native range *E. medicae* across most of their genome. This suggests that *Ensifer* genomes have more stable symbiotic replicons than other genera (Soberón-Chávez and Nájera 1989; Segovia, Pinero and Palacios 1991; Laguerre and Bardin 1993; Bromfield, Barran and Wheatcroft 1995; Sullivan

Table 1. Expected heterozygosity, private alleles, and FST calculated using alleles for single copy core gene SNPs within and between subpopulations.[a]

		Expected heterozygosity				Private alleles			FST
	# SNPs	CA	NAT	Overall	% Change in CA	CA	NAT	% Diff in CA	CA vs. NAT
Full genome	13 340	0.2844	0.1657	0.2533	+71.63*	5511	2445	+125.40	0.0983
Chromosome	6281	0.2520	0.1924	0.2366	+30.92*	2578	1288	+100.16	0.0671
psmed01	6397	0.3240	0.1346	0.2719	+140.77*	2742	881	+211.24	0.1301
psmed02	662	0.2099	0.2130	0.2317	− 1.44	191	276	− 30.80	0.0873

[a]Singleton alleles were excluded from these calculations to eliminate bias from potential sequencing errors. Calculations were made using all SNPs combined and those mapped to individual replicons in the WSM419 reference genome. No genes on psmed03 were core to the full set and thus we could not identify SNPs from this molecule. CA: California, Nat: native range. *Statistically significant differences between native and invaded range at $P < 0.0125$ (Bonferonni corrected cutoff) as determined by two-tailed *T*-test.

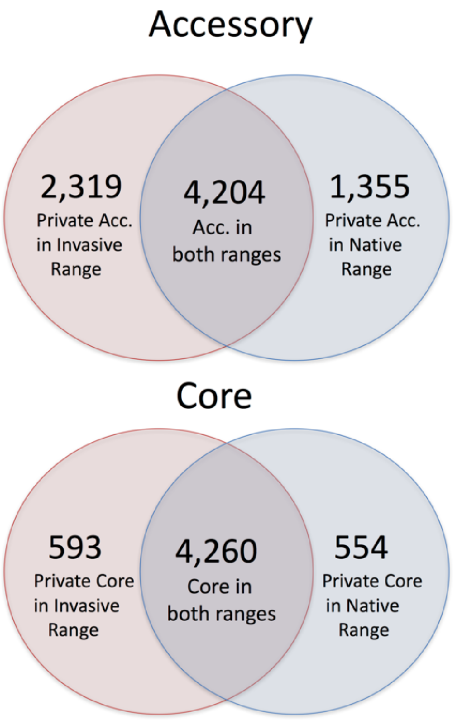

Figure 4. Comparison of *E. medicae* orthogroup attributes from pangenomes in the invaded vs. native range of the symbiosis. Accessory genes are those not found in all strains, while core genes are found to have at least one copy per strain. 'Private Accessory' is defined as orthogroups with at least one gene copy in one range but no copies in the other. 'Private Core' is defined as orthogroups with 100% allele frequency in one range but not the other.

Table 2. Orthologs at significantly higher frequencies in strains from either geographic range based on two-tailed *T*-test per orthogroup.

Significance threshold	Higher frequency in California invaded range	Higher frequency in native range
$P < 0.05$	1480	498
$P < 0.01$	761	162
Bonferroni corrected (<4.1E−6)	18	10

et al. 1996) and that differences in genome stability could underlie phylogenetically predictable patterns of microbial genome evolution during colonization.

Together, our data and historical records argue against a *M. polymorpha* invasion scenario via a preexisting cosmopolitan distribution of the symbiont in North America. First, while *E. medicae* nodulates diverse European legumes (Eardly *et al.* 1990), no native Californian legumes form nodules in nature with *E. medicae* (Lesins & Lesins 1979), and in the long-term absence of a host species, populations of rhizobia do not tend to persist (Selbitschka *et al.* 2006; Denison and Kiers 2011). Second, genomic diversity in the sampled portion of the invaded range is nested within native range diversity, consistent with introduction from Europe, and multiple European-origin *E. medicae* lineages co-occur in California. If *E. medicae* occurred in North America prior to European contact, the population should harbor unique subsets of deep diversity, due to a barrier to gene flow across the Atlantic Ocean. Third, historical records for the congener alfalfa (*Medicago sativa*), which nodulates with *E. medicae*, show North American soils initially lacked compatible rhizobia (Coburn 1907; Westgate 1908; Wing 1912). Thus, we conclude that *E. medicae* has co-invaded with *M. polymorpha*, possibly along with other compatible medic hosts. Co-invasions are a common pattern of legume-rhizobium invasion: *Mimosa* species co-invade with their beta-proteobacterial symbionts (Chen *et al.* 2005; Andrus, Andam and Parker 2012; Klonowska *et al.* 2012; Gehlot *et al.* 2013) and *Dipogon lignosus* co-invades with *Burkholderia* symbionts (Liu *et al.* 2014). Host-symbiont co-invasions may be common among diverse hosts as exemplified by the stinkbug, *Megacopta cribraria*, which has co-invaded with the nutritional symbiont, *Ishikawaella capulata* (Brown *et al.* 2014).

Population genomics of host-microbial symbiont co-invasion

Invaded range populations often exhibit low genetic diversity due to population bottlenecks at introduction (Dlugosch and Parker 2008; Barrett 2015; Bock *et al.* 2015). For example, an invasive mycorrhiza, *Amanita phalloides*, exhibits lower diversity and fewer private alleles in the portion of its range where it is exotic (Pringle *et al.* 2009). In contrast, we detected higher levels of diversity in invaded range *E. medicae* than in native range *E. medicae*, despite comparing a continent-wide native range *E. medicae* pangenome, to a pangenome sampled from a single reserve in the invaded range. Compared to *E. medicae* from *M. polymorpha*'s native range, the chromosome and psmed01 replicon from invaded range *E. medicae* exhibit higher heterozygosity at SNP loci and more private alleles. Much of this high diversity and high number of private alleles is driven by the psmed01 replicon, whereas the psmed02 replicon shows no significant difference in heterozygosity between ranges. This is surprising because psmed02, the replicon with the most symbiosis-related genes (Galardini *et al.* 2011; Sugawara *et al.* 2013), is considered a hotbed of recombination, gene gain and loss, and rapid evolution (Bailly *et al.* 2011), as is the psymA homolog in the sister species, *E. meliloti* (Giuntini *et al.* 2005). Our finding of increased nucleotide and private allele diversity in psmed01 and the chromosome, but not psmed02, in the Californian invaded range could result from selective forces acting on introduced variation, multiple introductions of highly diverse lineages from the native range, or *de novo* mutation. Purifying selection in native range *E. medicae* is thought to be weaker on the megaplasmids than on the chromosome (Epstein *et al.* 2012) possibly due to lower rates of plasmid gene expression (Morrow and Cooper 2012). Our findings suggest that different patterns of recombination and/or selection could underlie evolution in the native and invaded range.

While *E. medicae* harbors an order of magnitude less genotypic diversity than rhizobia such as *Rhizobium leguminosarum* (Epstein *et al.* 2012; Velázquez *et al.* 2015), using whole-genome analysis we were able to identify three clades of deep diversity that were present in our invaded range sample. We also found evidence for the lack of co-transmission across replicons, reflecting recombination or HGT, similar to that previously observed for *Ensifer* (Bailly *et al.* 2011, Epstein *et al.* 2012). *Ensifer medicae* nodulating another invasive medic, *Medicago lupulina*, in the Eastern USA contained even lower genetic diversity than we identified (Harrison *et al.* 2017). A longer and more mixed history of Spanish and English colonization (Spira and Wagner 1983; De Haan and Barnes 1998), combined with a Mediterranean climate could result in more diverse *E. medicae* in California than in

the Eastern USA. Alternatively, *M. lupulina* may associate with a smaller subset of *E. medicae* genotypes than does *M. polymorpha* (Brockwell and Hely 1966).

Novel microbial genomic content in the invaded range

The set of *E. medicae* strains from a single site in California harbor more accessory genes than those from sites spanning Europe and North Africa. Gene gain and loss play a major role in *E. medicae* population genomics (Bailly *et al.* 2011; Epstein *et al.* 2012; Sugawara *et al.* 2013; Epstein, Sadowsky and Tiffin 2014) and 65% of *E. medicae* orthologs were accessory genes in our study. California strains harbor more private accessory genes and have more accessory genes at higher frequency than the native range strains. While the majority of *E. medicae* accessory genes appear to be deleterious and subject to purifying selection in the native range (Epstein, Sadowsky and Tiffin 2014), novel conditions in California could drive selection for a greater number of adaptive accessory genes, or make *E. medicae* more vulnerable to the acquisition of deleterious accessory genes (Lee and Marx 2012; Epstein, Sadowsky and Tiffin 2014).

Compared to genes shared with native range strains, the California private accessory genome is enriched in loci with binding functions, which suggests that this molecular function could play a role in invasion. While most accessory genes share sequence identity with previously sequenced *Ensifer* strains, a subset do not and may have been acquired horizontally from other genera. Across the native range, private accessory genes are primarily enriched in transposon-related genes and other repetitive elements. The set of genes exclusive to Clade 1 strains in California are also enriched for transposon-related GO terms, including recombination, transposition and DNA metabolic processes. Clade 1 strains have the largest accessory genomes, indicating that gene gain and loss may play an especially large role in this lineage.

CONCLUSIONS

Our data support a host-symbiont co-invasion scenario for *M. polymorpha* and *E. medicae*, as opposed to invasion via novel microbial chimeras, novel microbial partner species or cosmopolitan generalist symbionts. Co-invading symbionts at a single reserve in California have greater genome-wide diversity than do symbionts from across the host's native range, suggesting no symbiont genetic bottleneck. In fact, compared to native range symbionts, invaded range symbionts harbor more accessory genes, which are uniquely enriched in binding molecular functions and could underlie novel adaptations. There is great potential for microbial genomics to shed light on symbiont invasion biology, which is a critical facet of the invasion biology of host plants and animals.

SUPPLEMENTARY DATA

Supplementary data are available at *FEMSEC* online.

ACKNOWLEDGEMENT

The authors thank J.P. Dunham for assistance with preparing genomic libraries, A. Garoutte for bioinformatics processing of raw sequence reads, B. Epstein for providing genomic information for some native range *E. medicae* strains, the authors of the publications on native range *E. medicae* genomes, and the editor and two anonymous reviewers for suggestions that improved the manuscript. The authors have no conflicts of interest to declare.

Data Accessibility: All sequence data are available under the NCBI BioProject PRJNA352431. The Whole Genome Shotgun sequences have been deposited at DDBJ/EMBL/GenBank under the accessions NBUA00000000-NBUQ00000000.

FUNDING

This work was funded by NSF DEB-1355216 to S.S. Porter and NSF DEB-1354878 to M.L. Friesen.

Conflict of interest. None declared.

REFERENCES

Alikhan N-F, Petty NK, Ben Zakour NL *et al.* BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* 2011;**12**:402.

Amsellem L, Brouat C, Duron O *et al.* Importance of microorganisms to macroorganisms invasions: is the essential invisible to the eye?(The Little Prince, A. de Saint-Exupéry, 1943). *Adv Ecol Res* 2016;**57**:99–146.

Ana Conesa SG. Blast2GO: a comprehensive suite for functional analysis in plant genomics. *Int J Plant Gen* 2008;**2008**:1–12.

Andrus AD, Andam C, Parker MA. American origin of *Cupriavidus* bacteria associated with invasive *Mimosa* legumes in the Philippines. *FEMS Microb Ecol* 2012;**80**:747–50.

Aziz RK, Bartels D, Best AA *et al.* The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* 2008;**9**:75.

Bailly X, Giuntini E, Sexton MC *et al.* Population genomics of *Sinorhizobium medicae* based on low-coverage sequencing of sympatric isolates. *ISME J* 2011;**5**:1722–34.

Bailly X, Olivieri I, De Mita S *et al.* Recombination and selection shape the molecular diversity pattern of nitrogen-fixing *Sinorhizobium sp.* associated to *Medicago*. *Mol Ecol* 2006;**15**:2719–34.

Barrett SCH. Foundations of invasion genetics: the Baker and Stebbins legacy. *Mol Ecol* 2015;**24**:1927–41.

Bock DG, Caseys C, Cousens RD *et al.* What we still don't know about invasion genetics. *Mol Ecol* 2015;**24**:2277–97.

Bogar LM, Dickie IA, Kennedy PG. Testing the co-invasion hypothesis: ectomycorrhizal fungal communities on *Alnus glutinosa* and *Salix fragilis* in New Zealand. *Diversity Distrib* 2015;**21**:268–78.

Bonito G, Trappe JM, Donovan S *et al.* The Asian black truffle *Tuber indicum* can form ectomycorrhizas with North American host plants and complete its life cycle in non-native soils. *Fungal Ecol* 2011;**4**:83–93.

Brockwell J, Hely FW. Symbiotic characteristics of *Rhizobium meliloti*: and appraisal of the systematic treatment of nodulation and nitrogen fixation interactions between hosts and rhizobia of diverse origins. *Aust J Agric Res* 1966;**17**:885–99.

Bromfield ESP, Barran LR, Wheatcroft R. Relative genetic structure of a population of *Rhizobium meliloti* isolated directly from soil and from nodules of alfalfa (*Medicago sativa*) and sweet clover (*Melilotus alba*). *Mol Ecol* 1995;**4**:183–8.

Brown AMV, Huynh LY, Bolender CM *et al.* Population genomics of a symbiont in the early stages of a pest invasion. *Mol Ecol* 2014;**23**:1516–30.

Chen WM, James EK, Chou JH *et al.* Beta-rhizobia from *Mimosa pigra*, a newly discovered plant in Taiwan. *New Phyt* 2005;**168**:661–75.

Coburn FD. *The Book of Alfalfa*. Orange Judd Company, 1907.

Crisóstomo JA, Rodríguez-Echeverría S, Freitas H. Co-introduction of exotic rhizobia to the rhizosphere of the invasive legume *Acacia saligna*, an intercontinental study. *Appl Soil Ecol* 2013;**64**:118–26.

Daehler CC. The taxonomic distribution of invasive angiosperm plants: ecological insights and comparison to agricultural weeds. *Biol Conserv* 1998;**84**:167–80.

Darmon E, Leach DRF. Bacterial genome instability. *Microbiol Mol Biol Rev* 2014;**78**:1–39.

De Haan RL, Barnes DK. Inheritance of pod type, stem color, and dwarf growth habit in *Medicago polymorpha*. *Crop Sci* 1998;**38**:1558–61.

De Wit R, Bouvier T. "Everything is everywhere, but, the environment selects;" what did Baas Becking and Beijerinck really say? *Environ Microbiol* 2006;**8**:755–8.

Denison RF, Kiers ET. Life histories of symbiotic rhizobia and mycorrhizal fungi. *Curr Biol* 2011;**21**:R775–85.

Desprez-Loustau M-L, Robin C, Buée M *et al*. The fungal dimension of biological invasions. *Trends Ecol Evol* 2007;**22**:472–80.

Dickie IA, Bolstridge N, Cooper JA *et al*. Co-invasion by *Pinus* and its mycorrhizal fungi. *New Phytol* 2010;**187**:475–84.

Dlugosch KM, Parker IM. Founding events in species invasions: genetic variation, adaptive evolution, and the role of multiple introductions. *Mol Ecol* 2008;**17**:431–49.

Dunham JP, Friesen ML. A cost-effective method for high-throughput construction of Illumina sequencing libraries. *Cold Spring Harb Protoc* 2013;**2013**:820–34.

Eardly BD, Materon LA, Smith NH *et al*. Genetic structure of natural populations of the nitrogen-fixing bacterium *Rhizobium meliloti*. *Appl Environ Microbiol* 1990;**56**:187–94.

Epstein B, Branca A, Mudge J *et al*. Population genomics of the facultatively mutualistic bacteria *Sinorhizobium meliloti* and *S. medicae*. PLoS Genet 2012;**8**:e1002868.

Epstein B, Sadowsky MJ, Tiffin P. Selection on horizontally transferred and duplicated genes in *Sinorhizobium* (*Ensifer*), the root-nodule symbionts of *Medicago*. *Genome Biol Evol* 2014;**6**:1199–209.

Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol Ecol* 2005;**14**:2611–20.

Finlay BJ. Global dispersal of free-living microbial eukaryote species. *Science* 2002;**296**:1061–3.

Friesen ML, Porter SS, Stark SC *et al*. Microbially mediated plant functional traits. *Annu Rev Ecol Evol Syst* 2011;**42**:23–46.

Galardini M, Mengoni A, Brilli M *et al*. Exploring the symbiotic pangenome of the nitrogen-fixing bacterium *Sinorhizobium meliloti*. *BMC Genomics* 2011;**12**:235.

Galibert F, Finan TM, Long SR *et al*. The composite genome of the legume symbiont *Sinorhizobium meliloti*. *Science* 2001;**293**:668–72.

Gehlot HS, Tak N, Kaushik M *et al*. An invasive *Mimosa* in India does not adopt the symbionts of its native relatives. *Annal Bot* 2013;**112**:179–96.

Giuntini E, Mengoni A, De Filippo C *et al*. Large-scale genetic variation of the symbiosis-required megaplasmid pSymA revealed by comparative genomic analysis of *Sinorhizobium meliloti* natural strains. *BMC Genomics* 2005;**6**:158.

Hanson CA, Fuhrman JA, Horner-Devine MC *et al*. Beyond biogeographic patterns: processes shaping the microbial landscape. *Nat Rev Microbiol* 2012;**10**:497–506.

Harrison TL, Wood CW, Heath KD *et al*. Geographically structured genetic variation in the *Medicago lupulina- Ensifer* mutualism. *Evolution* 2017;**71**:1787–801.

Haskett TL, Terpolilli JJ, Bekuma A *et al*. Assembly and transfer of tripartite integrative and conjugative genetic elements. *P Natl Acad Sci USA* 2016;**113**:12268–73.

Hayashi M, Miyahara A, Sato S *et al*. Construction of a genetic linkage map of the model legume *Lotus japonicus* using an intraspecific F-2 population. *DNA Res* 2001;**8**:301–10.

Hayward J, Horton TR, Nuñez MA. Ectomycorrhizal fungal communities coinvading with *Pinaceae* host plants in Argentina: Gringos bajo el bosque. *New Phytol* 2015;**208**:497–506.

Hayward J, Horton TR, Pauchard A *et al*. A single ectomycorrhizal fungal species can enable a *Pinus* invasion. *Ecology* 2015;**96**:1438–44.

Himler AG, Adachi-Hagimori T, Bergen JE *et al*. Rapid spread of a bacterial symbiont in an invasive whitefly is driven by fitness benefits and female bias. *Science* 2011;**332**:254–6.

Horn K, Parker IM, Malek W *et al*. Disparate origins of *Bradyrhizobium* symbionts for invasive populations of *Cytisus scoparius* (*Leguminosae*) in North America. *FEMS Microbiol Ecol* 2014;**89**:89–98.

Hubisz MJ, Falush D, Stephens M *et al*. Inferring weak population structure with the assistance of sample group information. *Mol Ecol Resour* 2009;**9**:1322–32.

Huson DH. SplitsTree: analyzing and visualizing evolutionary data. *Bioinformatics* 1998;**14**:68–73.

Inderjit, van der Putten WH. Impacts of soil microbial communities on exotic plant invasions. *Trends Ecol Evol* 2010;**25**:512–9.

Klonowska A, Chaintreuil C, Tisseyre P *et al*. Biodiversity of *Mimosa pudica* rhizobial symbionts (*Cupriavidus taiwanensis, Rhizobium mesoamericanum*) in New Caledonia and their adaptation to heavy metal-rich soils. *FEMS Microbiol Ecol* 2012;**81**:618–35.

Kuo CH, Ochman H. The fate of new bacterial genes. *Fems Microbiol Rev* 2009;**33**:38–43.

Kück P, Meusemann K. FASconCAT: convenient handling of data matrices. *Mol Phylogenet Evol* 2010;**56**:1115–8.

Lachmuth S, Durka W, Schurr FM. The making of a rapid plant invader: genetic diversity and differentiation in the native and invaded range of *Senecio inaequidens*. *Mol Ecol* 2010;**19**:3952–67.

Lafay B, Burdon JJ. Molecular diversity of rhizobia nodulating the invasive legume *Cytisus scoparius* in Australia. *J Appl Microbiol* 2006;**100**:1228–38.

Laguerre G, Bardin M. Isolation from soil of symbiotic and non-symbiotic *Rhizobium leguminosarum* by DNA hybridization. *Can J Microbiol* 1993;**12**:1142–9.

Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Meth* 2012;**9**:357–9.

LaPierre K, Simms E, Tariq M *et al*. Invasive legumes can associate with many mutualists of native legumes, but usually don't. *Ecol Evol* 2017;**7**:8599–611.

Le Roux JJ, Hui C, Keet J-H *et al*. Co-introduction vs ecological fitting as pathways to the establishment of effective mutualisms during biological invasions. *New Phytol* 2017;**400**:193.

Lee MC, Marx CJ. Repeated, selection-driven genome reduction of accessory genes in experimental populations. *PLoS Genet* 2012;**8**:e1002651.

Lesins KA, Lesins I. *Genus Medicago (Leguminosae): A Taxogenetic Study*. Hingham, Boston, MA: Kluwer, 1979.

Liu WYY, Ridgway HJ, James TK *et al*. *Burkholderia* sp. induces functional nodules on the South African invasive legume *Dipogon lignosus* (Phaseoleae) in New Zealand soils. *Microb Ecol* 2014;**68**:542–55.

Lischer HE, Excoffier L. PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics* 2012;**28**:298–9.

Litchman E. Invisible invaders: non-pathogenic invasive microbes in aquatic and terrestrial ecosystems. *Ecol Lett* 2010;**13**:1560–72.

Lowry E, Rollinson EJ, Laybourn AJ *et al.* Biological invasions: a field synopsis, systematic review, and database of the literature. *Ecol Evol* 2013;**3**:182–96.

Mack RN, Simberloff D, Mark Lonsdale W *et al.* Biotic invasions: causes, epidemiology, global consequences, and control. *Ecol Appl* 2000;**10**:689–710.

Marquez LM, Redman RS, Rodriguez RJ *et al.* A virus in a fungus in a plant: three-way symbiosis required for thermal tolerance. *Science* 2007;**315**:513–5.

Moora M, Berger S, Davison J *et al.* Alien plants associate with widespread generalist arbuscular mycorrhizal fungal taxa: evidence from a continental-scale study using massively parallel 454 sequencing. *J Biogeogr* 2011;**38**:1305–17.

Morrow JD, Cooper VS. Evolutionary effects of translocations in bacterial genomes. *Genome Biol Evol* 2012;**4**:1256–62.

Nandasena KG, O'Hara GW, Tiwari RP *et al.* In situ lateral transfer of symbiosis islands results in rapid evolution of diverse competitive strains of mesorhizobia suboptimal in symbiotic nitrogen fixation on the pasture legume *Biserrula pelecinus* L. *Environ Microbiol* 2007;**9**:2496–511.

Narożna D, Pudełko K, Króliczak J *et al.* Survival and competitiveness of *Bradyrhizobium japonicum* strains 20 years after introduction into field locations in Poland. *Appl Environ Microbiol* 2015;**81**:5552–9.

Nguyen DT, Spooner-Hart RN, Riegler M. Loss of *Wolbachia* but not *Cardinium* in the invasive range of the Australian thrips species, *Pezothrips kellyanus*. *Biol Invasions* 2016;**18**:197–214.

Nougué O, Gallet R, Chevin L-M *et al.* Niche limits of symbiotic gut microbiota constrain the salinity tolerance of brine shrimp. *Am Nat* 2015;**186**:390–403.

Nuñez MA, Dickie IA. Invasive below ground mutualists of woody plants. *Biol Invasions* 2014;**16**:645–61.

Nuñez MA, Horton TR, Simberloff D. Lack of belowground mutualisms hinders *Pinaceae* invasions. *Ecology* 2009;**90**: 2352–9.

Ochman H, Lawrence JG, Groisman EA. Lateral gene transfer and the nature of bacterial innovation. *Nature* 2000;**405**:299–304.

Oldroyd GED. Speak, friend, and enter: signalling systems that promote beneficial symbiotic associations in plants. *Nat Rev Microbiol* 2013;**11**:252–63.

Oliver KM, Degnan PH, Burke GR *et al.* Facultative symbionts in aphids and the horizontal transfer of ecologically important traits. *Annu Rev Entomol* 2010;**55**:247–66.

Parker IM, Gilbert GS. The evolutionary ecology of novel plant-pathogen interactions. *Annu Rev Ecol Evol Syst* 2004;**35**:675–700.

Parker MA, Malek W, Parker IM. Growth of an invasive legume is symbiont limited in newly occupied habitats. *Diversity Distrib* 2006;**12**:563–71.

Parker MA, Wurtz AK, Paynter Q. Nodule symbiosis of invasive *Mimosa pigra* in Australia and in ancestral habitats: a comparative analysis. *Biol Invasions* 2006;**9**:127–38.

Parker MA. Mutualism as a constraint on invasion successs for legumes and rhizobia. *Diversity Distrib* 2001;**7**: 125–36.

Pimentel D, McNair S, Janecka J *et al.* Economic and environmental threats of alien plant, animal, and microbe invasions. *Agric Ecosyst Environ* 2001;**84**:1–20.

Pimentel D ed. *Biological Invasions: Economic and Environmental Costs of Alien Plant, Animal, and Microbe Species*, 2nd edn. USA: CRC Press, 2011.

Polz MF, Alm EJ, Hanage WP. Horizontal gene transfer and the evolution of bacterial and archaeal population structure. *Trends Genet* 2013;**29**:170–5.

Porter SS, Chang PL, Conow CA *et al.* Association mapping reveals novel serpentine adaptation gene clusters in a population of symbiotic *Mesorhizobium*. *ISME J* 2016;**11**:248.

Porter SS, Rice KJ. Trade-offs, spatial heterogeneity, and the maintenance of microbial diversity. *Evolution* 2013;**67**:599–608.

Pringle A, Bever JD, Gardes M *et al.* Mycorrhizal symbioses and plant invasions. *Annu Rev Ecol Evol Syst* 2009;**40**:699–715.

Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics* 2000;**155**:945–59.

Puzey J, Vallejo-Marín M. Genomics of invasion: diversity and selection in introduced populations of monkeyflowers (*Mimulus guttatus*). *Mol Ecol* 2014;**23**:4472–85.

Raymond M, Rousset F. GENEPOP (Version 1.2): population genetics software for exact tests and ecumenicism. *J Hered* 1995;**83**:239.

Reeve W, Ardley J, Tian R *et al.* A genomic encyclopedia of the root nodule bacteria: assessing genetic diversity through a systematic biogeographic survey. *Stand Genomic Sci* 2015;**10**:14.

Reeve W, Chain P, O'Hara G. Complete genome sequence of the *Medicago* microsymbiont *Ensifer* (*Sinorhizobium*) *medicae* strain WSM419. *Stand Genomic Sci* 2010;**2**:77–86.

Rice P, Longden I, Bleasby A. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet* 2000;**16**:276–7.

Richardson DM, Allsopp N, D'Antonio CM *et al.* Plant invasions—the role of mutualisms. *Biol Rev* 2000;**75**:65–93.

Rome S, Fernandez MP, Brunel B *et al. Sinorhizobium medicae sp.* nov., isolated from annual *Medicago spp*. *Int J Syst Bacteriol* 1996;**46**:972–80.

Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 2011;**27**:863–4.

Segovia L, Pinero D, Palacios R. Genetic structure of a soil population of nonsymbiotic *Rhizobium leguminosarum*. *Appl Environ Microbiol* 1991;**57**:426–33.

Selbitschka W, Keller M, Miethling-Graff R *et al.* Long-term field release of bioluminescent *Sinorhizobium meliloti* strains to assess the influence of a recA mutation on the strains' survival. *Microb Ecol* 2006;**52**:583–95.

Silva C, Kan FL, Martinez-Romero E. Population genetic structure of *Sinorhizobium meliloti* and *S. medicae* isolated from nodules of *Medicago spp*. in Mexico. *FEMS Microbiol Ecol* 2007;**60**:477–89.

Silva C, Vinuesa P, Eguiarte LE *et al.* Evolutionary genetics and biogeographic structure of *Rhizobium gallicum sensu lato*, a widely distributed bacterial symbiont of diverse legumes. *Mol Ecol* 2005;**14**:4033–50.

Simonsen AK, Dinnage R, Barrett LG *et al.* Symbiosis limits establishment of legumes outside their native range at a global scale. *Nat Commun* 2017;**8**, DOI: 10.1038/ncomms14790.

Soberón-Chávez G, Nájera R. Isolation from soil of *Rhizobium leguminosarum* lacking symbiotic information. *Can J Microbiol* 1989;**35**:464–8.

Spira TP, Wagner LK. Viability of seeds up to 211 years old extracted from adobe brick buildings of California and Northern Mexico. *Am J Bot* 1983;**70**:303–7.

Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 2014;**30**:1312–3.

Stanton-Geddes J, Anderson CG. Does a facultative mutualism limit species range expansion? *Oecologia* 2011;**167**:149–55.

Strayer DL. Eight questions about invasions and ecosystem functioning *Ecol Lett* 2012;**15**:1199–1210.

Sugawara M, Epstein B, Badgley BD *et al*. Comparative genomics of the core and accessory genomes of 48 *Sinorhizobium* strains comprising five genospecies. *Genome Biol* 2013;**14**:R17.

Sullivan JT, Eardly BD, van Berkum P *et al*. Four unnamed species of nonsymbiotic rhizobia isolated from the rhizosphere of *Lotus corniculatus*. *Appl Environ Microbiol* 1996;**62**:2818–25.

Sullivan JT, Patrick HN, Lowther WL *et al*. Nodulating strains of *Rhizobium loti* arise through chromosomal symbiotic gene transfer in the environment. *P Natl Acad Sci USA* 1995;**92**:8985–9.

Sullivan JT, Ronson CW. Evolution of rhizobia by acquisition of a 500-kb symbiosis island that integrates into a phe-tRNA gene. *P Natl Acad Sci USA* 1998;**95**:5145–9.

Traveset A, Richardson DM. Mutualistic interactions and biological invasions. *Annu Rev Ecol Evol Syst* 2014;**45**:89–113.

Tritt A, Eisen JA, Facciotti MT *et al*. An integrated pipeline for de novo assembly of microbial genomes. *PLoS ONE* 2012;**7**:e42304.

van der Putten WH, Klironomos JN, Wardle DA. Microbial ecology of biological invasions. *ISME J* 2007;**1**:28–37.

Velázquez E, Peix A, Zurdo-Piñiro JL *et al*. The coexistence of symbiosis and pathogenicity-determining genes in *Rhizobium rhizogenes* strains enables them to induce nodules and tumors or hairy roots in plants. *Open Biol* 2015;**18**:140133–3.

Vilcinskas A, Stoecker K, Schmidtberg H *et al*. Invasive harlequin ladybird carries biological weapons against native competitors. *Science* 2013;**340**:862–3.

Vitousek PM, D'Antonio CM, Loope LL. Introduced species: a significant component of human-caused global change. *New Zeal J Ecol* 1997:1–16.

Wang S, Hao B, Li J *et al*. Whole-genome sequencing of *Mesorhizobium huakuii* 7653R provides molecular insights into host specificity and symbiosis island dynamics. *BMC Genomics* 2003;**13**:2178–89.

Westgate JM. *Farmers' Bulletin 339. Alfalfa*. Washington, DC: United States Department of Agriculture, 1908.

Wing JE. *Alfalfa Farming in America*. Chicago, IL: Sanders Publishing Company, 1912.

Wolfe BE, Richard F, Cross HB *et al*. Distribution and abundance of the introduced ectomycorrhizal fungus *Amanita phalloides* in North America. *New Phytol* 2010;**185**:803–16.

Young JPW. Bacteria are smartphones and mobile genes Are apps. *Trends Microbiol* 2016;**24**:931–2.

Zhang Y, Zhang D, Barrett S. Genetic uniformity characterizes the invasive spread of water hyacinth (*Eichhornia crassipes*), a clonal aquatic plant. *Mol Ecol* 2010;**19**:1774–86.