MOLECULAR ECOLOGY

Molecular Ecology (2013) doi: 10.1111/mec.12366

SPECIAL ISSUE: NATURE'S MICROBIOME

Population genomics of a symbiont in the early stages of a pest invasion

AMANDA M. V. BROWN,* LYNN Y. HUYNH,† CAITLIN M. BOLENDER,* KELLY G. NELSON* and JOHN P. MCCUTCHEON*

*Division of Biological Sciences, University of Montana, 32 Campus Drive, HS104, Missoula, MT 59812, USA, †Department of Biology, Emory University, 1510 Clifton Road NE, Atlanta, GA 30322, USA

Abstract

Invasive species often depend on microbial symbionts, but few studies have examined the evolutionary dynamics of symbionts during the early stages of an invasion. The insect *Megacopta cribraria* and its bacterial nutritional symbiont *Candidatus* Ishikawaella capsulata invaded the southeastern US in 2009. While *M. cribraria* was initially discovered on wild kudzu plants, it was found as a pest on soybeans within 1 year of infestation. Because prior research suggests *Ishikawaella* confers the pest status—that is, the ability to thrive on soybeans—in some *Megacopta* species, we performed a genomic study on *Ishikawaella* from US. *Megacopta cribraria* populations to understand the role of the symbiont in driving host plant preferences. We included *Ishikawaella* samples collected in the first days of the invasion in 2009 and from 23 locations across the insect's 2011 US range. The 0.75 Mb symbiont genome revealed only 47 fixed differences from the pest-conferring *Ishikawaella* in Japan, with only one amino acid change in a nutrition-provisioning gene. This similarity, along with a lack of fixed substitutions in the US symbiont population, indicates that *Ishikawella* likely arrived in the US capable of being a soybean pest. Analyses of allele frequency changes between 2009 and 2011 uncover signatures of both positive and negative selection and suggest that symbionts on soybeans and kudzu experience differential selection for genes related to nutrient provisioning. Our data reveal the evolutionary trajectory of an important insect-bacteria symbiosis in the early stages of an invasion, highlighting the role microbial symbionts may play in the spread of invasive species.

Keywords: invasive species, *Ishikawaella*, *Megacopta*, population genomics, soybean pest, symbiont

Received 16 December 2012; revision received 6 March 2013; accepted 14 March 2013

Introduction

Microbial symbionts often accompany invading species (Desprez-Loustau *et al.* 2007; Janson *et al.* 2008; Lichman 2010; Feldhaar 2011) and could, in theory, play a role in the biology of invasions. By providing a limiting resource, nutrient-provisioning symbionts are perhaps especially likely to exert this control on their hosts (Clay & Holah 1999). Understanding the role that symbionts play in the establishment and maintenance of species invasions could therefore be helpful in management decisions. Despite the potential importance of this phenomenon, few studies have addressed microbial symbiont evolution during the early stages of invasion (Shah *et al.* 2009; Feldhaar 2011). Sap-feeding insects within the order Hemiptera have the potential to show this symbiont-driven evolution, as many of these insects have nutrient-poor diets, depend on obligate bacterial symbionts to synthesize amino acids or vitamins absent in their food, and display sometimes extreme ecological specialization (Buchner 1965; Baumann 2005; Dale & Moran 2006; McCutcheon & Moran 2010; MacDonald *et al.* 2011; Sachs *et al.* 2011; Vogel & Moran 2011).

Correspondence: Amanda M. V. Brown, Fax: 406 243 4184; E-mail: amanda.brown@mso.umt.edu

However, as many symbionts in this group show strictly vertical transmission, undergo little or no recombination, and have extremely reduced genomes with no evidence of horizontal gene transfer (Mira *et al.* 2001; Andersson 2008; McCutcheon & Moran 2011), it is likely that they have a reduced capacity for adaptation compared with free-living bacteria. It is therefore unclear whether these symbionts could provide much dynamic genetic potential when placed in a new environment. Nevertheless, many Hemipterans are highly successful invasive pests (e.g. whiteflies, aphids, scale insects, psyllids and stinkbugs; www.issg.org; Pimentel *et al.* 2005), and the role of symbionts in these invasions is unknown.

Here, we study a recent invasive insect-bacteria symbiosis whose symbiont governs its success on different host plants (Hosokawa *et al.* 2007). *Candidatus* Ishikawaella capsulata (hereafter, *Ishikawaella*; Gammaproteobacteria) is a symbiont of stinkbugs in the genus *Megacopta* (Hemiptera: Plataspidae). *Ishikawaella* lives in specialized crypts in the insect's midgut, synthesizing essential amino acids and vitamins missing in the phloem sap diet of the insect (Fukatsu & Hosokawa 2002; Hosokawa *et al.* 2005, 2006). These insects are legume (Fabaceae) specialists endemic to Asia and previously unknown in the New World (Eger *et al.* 2010). However, *Megacopta cribraria* was discovered in Hall Co., Georgia in October 2009 and has since spread rapidly to eight states in the southeastern US covering more than 400 000 km^2, following the range of its preferred wild host, kudzu (*Pueraria* spp.; Eger *et al.* 2010; Suiter *et al.* 2010; Zhang *et al.* 2012). Within 9 months of the initial US invasion, *M. cribraria* was found on soybeans (*Glycine max*) (Suiter *et al.* 2010) and is now a widespread US soybean crop pest. This was unexpected because the US insect genetically and morphologically resembles *M. cribraria* in Asia, which are considered to be 'nonpests' that rarely infest crop legumes (Hosokawa *et al.* 2005, 2007; Jenkins & Eaton 2011). In Japan, a closely related but genetically distinct insect (*Megacopta punctatissima*) is a frequent pest of soybeans, having significantly higher fitness on soybeans than the nonpest *M. cribraria* (Hosokawa *et al.* 2007). Remarkably, Hosokawa *et al.* (2007) showed that by exchanging the symbiont-containing transmission 'capsules' that females lay down with eggs, they could reverse the pest and nonpest phenotypes of the insects (Fig. 1A), effectively making the nonpest into a pest (Hosokawa *et al.* 2007). They concluded it is the genotype of the symbiont *Ishikawaella* rather than that of the host that determines the pest status.

Did the symbiosis arrive in the US with a nonpest *Ishikawella* genotype and then evolve to be a pest of soybeans or was the founding *Ishikawella* genotype capable of being a pest from the beginning? While these questions are complicated by a lack information on the role of the insect in the pest status of the symbiosis, understanding the early evolutionary trajectory of the symbi-

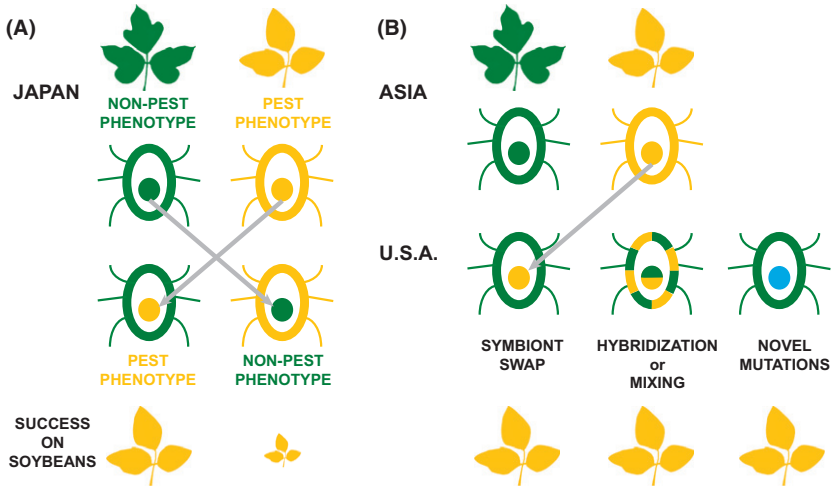

Fig. 1 Experimental symbiont swap by Hosokawa *et al.* (2007) and alternate scenarios for emergence of the pest phenotype in the US invasion. (A) Results of experimental symbiont swap by Hosokawa *et al.* (2007) showing that nonpest and pest insects (*Megacopta cribraria* and *Megacopta punctatissima*, respectively) reversed their differential hatch rate on soybeans (represented by different soybean leaf size at bottom). While both *Megacopta* spp. thrive on kudzu (shown in green), only the pest thrives on soybeans (yellow). This demonstrated that the genotype of the symbiont *Ishikawaella* determines the insect phenotype. (B) Alternate scenarios for emergence of the pest phenotype on US soybeans (yellow) from a nonpest insect in Asia (green): horizontal acquisition of the symbiont from a pest insect (i.e. 'symbiont swap') before arrival in the US, mixing or hybridization amongst pest and nonpest insects and/or recombination in symbionts in Asia, or novel 'pest' mutations in the symbiont in the US (blue).

ont is of great practical interest, as infestation with this insect-bacteria symbiotic system significantly decreases growth of both plants. Soybeans are worth approximately $40 billion in the US, and kudzu is a significant invasive plant (Kikuchi & Kobayashi 2010; Zhang *et al.* 2012; www.ers.usda.gov/topics/crops/soybeans-oil-crops.aspx). For nutritional symbionts, it is easy to envision small genetic changes causing large effects. For example, substitutions or indels could occur in pathways essential to the symbiont's nutrition-provisioning role. If plants differ in the availability of nutrients synthesized by the symbiont, these genetic changes could affect the insect's success on different plants. Such changes could be deleterious, such as nonsense or frameshift mutations disrupting symbiont nutrient pathway genes, or they could be beneficial if they increase production of a limiting nutrient. The recent publication of the completed genome of *Ishikawaella* from the pest insect *M. punctatissima* (Nikoh *et al.* 2011) simplifies our search for genomic differences that may be associated with *M. cribraria*'s emergence on soybeans in the US.

We used population genomics to explore the evolutionary trajectory of *Ishikawaella* during the first 2 years of the US invasion. By deep sequencing the complete *Ishikawaella* genome from samples across the US—including a sample from the first days of infestation (hereafter referred to as 'ground zero')—the current study tests various scenarios for pest emergence during the invasion (Fig. 1B). We were particularly interested in determining the similarity of the US symbiont to the pest-conferring symbiont in Asia, as well as the magnitude and direction of genetic change that has taken place since the founding of the invasion. This is important as the insects display a mixed pattern of soybean infestation in the US, with some localities not as affected as others (Suiter *et al.* 2010; Jenkins & Eaton 2011), and colour and size variation in the insects, with at least one case apparently varying with host plant (see Fig. S1, Supporting information). Insect size and colour are indicative of symbiont genotype in Asia, the nonpest being smaller and paler in colour than the pest, and symbiont-free (aposymbiotic) insects being even more small and pale (Hosokawa *et al.* 2006, 2007). Do these variations in plant preference and insect morphology in the US correlate with symbiont genetic diversity?

Our results show that the symbiont genotype has not changed during its time in the US, including during the emergence on soybeans early in its invasion. We show this genotype strongly resembles the pest-conferring *Ishikawaella* genotype in *M. punctatissima* in Japan and hypothesize that the symbiosis arrived in the US capable of being a pest and has maintained this ability from 2009 to 2011. At the same time, our analyses identify genes of potential functional interest for the pest status in this system.

Methods

Sample collection

Megacopta cribraria were collected in the US (Table S1, Supporting information) from 15 July to 8 October 2011 from wild kudzu vine (*Pueraria montana* var. *lobata*), bush clover (*Lespedeza cuneata*) and cultivated soybeans (*Glycine max*) at locations across the insect's range (Fig. 2). These insects were frozen at −20 or −80 °C. A specimen was also collected during the first week of observed infestation in the US in Hall Co., Georgia (or 'ground zero'), on 3 November 2009 from *Pueraria* spp. and preserved in 95% EtOH. It should be clearly noted that our ground zero sample came from a single insect and that this insect was stored differently than the remaining samples. This was primarily due to necessity—very few insects from the first sightings were preserved using methods amenable to DNA work. It is therefore possible that our ground zero polymorphism frequencies are skewed in some way, but the available evidence is consistent with the outbreak beginning with a very small number of individuals (Eger *et al.* 2010; Jenkins & Eaton 2011), and we are not aware of any systematic bias that would result from storage in ethanol vs. freezing.

Sample preparation and quantitative PCR

Insects were dissected in sterile Petri plates in phosphate-buffered saline with Tween 20 (PBST). Forceps and dissecting scissors were scrubbed with detergent, rinsed with 50% bleach and wiped with ethanol between dissections. The symbiont-crypt bearing gut section was removed, washed in fresh PBST, placed in a 1.5 ml tube and ground with a micropestle in 20 μl proteinase K and 180 μl ATL buffer followed by overnight digestion and DNA isolation using DNeasy Blood and Tissue Kit (QIAGEN) following the manufacturer's directions. DNA was quantitated for each sample in triplicate using the QuantiFluor dsDNA System (Promega) on a STRATAGENE MX3000P by a standard curve assay. To assess quality of dissections and optimize coverage of the symbiont in sequencing, quantitative PCR was performed using a STRATAGENE MX3000P with Brilliant III Ultra-Fast SYBR Green QPCR Master Mix (Agilent) and *Ishikawaella*-specific groEL primers (Hosokawa *et al.* 2007) normalized using the insect elongation factor 1-alpha gene, with 95 °C 3 m, 40 cycles 95 °C 20 s, 60 °C 20 s and a final melt of 95 °C 30 s, 55 °C 30 s, 72 °C 1 m. One genomic library was

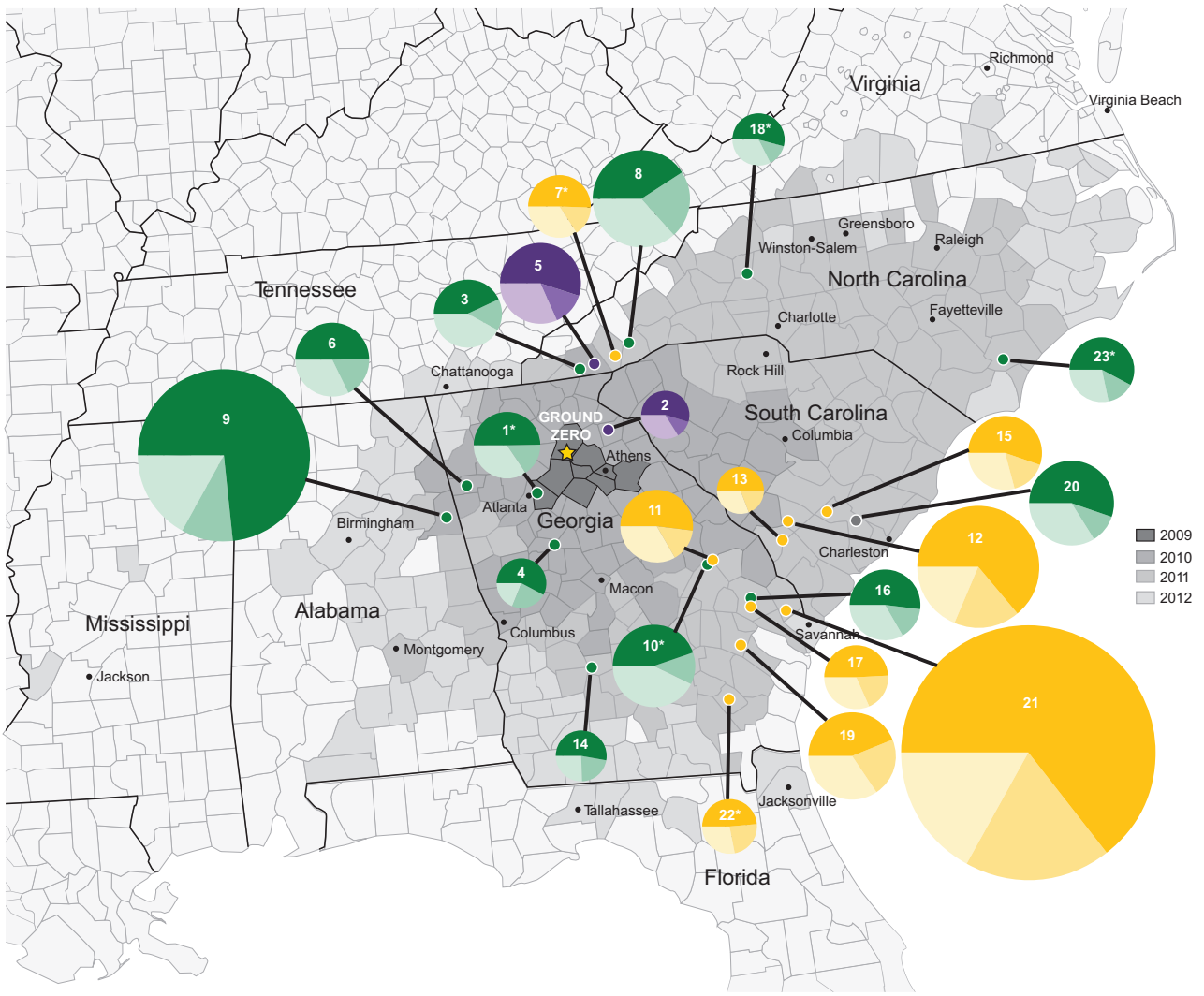

Fig. 2 Map showing US *Megacopta cribraria* range (2009–2012) and collection locations used for sequencing, with pie graphs showing polymorphic allele frequency change in *Ishikawaella capsulata* since the origin of invasion in Hall Co. Georgia, Oct. 2009 (or 'ground zero'). Pie graph sizes indicate absolute allele frequency change since ground zero for each sample. Pie graph wedges indicate proportion of nonsynonymous, synonymous and noncoding change (dark, middle and light shades, respectively). Green = kudzu, Yellow = soybeans, Purple = *Lespedeza*. Grey shading from darkest to lightest represents the insect's range from 2009 to 2012. Numbers correspond to sample locations in order of distance from ground zero. Samples marked with asterisks consisted of 15 insects.

prepared for each sample location (Fig. 2). For 18 locations, samples comprised single insects and for six locations, samples comprised 15 insects. In all cases, we chose specimens with the highest symbiont-to-insect ratio. This resulted in 24 libraries comprised of 108 insects.

Genomic library creation and assembly

Genomic libraries were prepared based on protocols described in Son & Taylor (2011) and instructions supplied with the NEXTflex DNA Barcodes (Bioo Scientific, Austin, TX; see Data S1, Supporting information). Barcoded libraries were demultiplexed, quality filtered and assembled using standard protocols, described in detail in Data S1, Supporting information. The US *Ishikawaella* genome was represented at an average of 266× depth per insect (N = 108 insects) for a combined coverage of 1196× per sampling location (for details see Table S2, Supporting information). The *Ishikawaella* plasmid always assembled into a single fragment with negligible polymorphism. It was sequenced at, on average, approximately 3500 × coverage per sampling location, corresponding to 3 copies per *Ishikawaella* genome. To rule out the possibility of other microbial symbionts

that may be contributing to the symbiosis in some way, we performed BLASTX searches against the GenBank non-redundant protein database with all assembled contigs not assigned to *Ishikawella* and the insect mitochondria.

Mapping, SNP analysis and annotation

Consensus sequences from each *de novo* assembly were used as references for mapping and SNP/indel calling. GATK v.1.5-16 (McKenna *et al.* 2010; DePristo *et al.* 2011) was used for local realignment. The pipeline consisted of using paired-end mapping with bwa and SAMTOOLS (Li & Durbin 2009; Li *et al.* 2009), followed by local realignment using RealignerTargetCreator and IndelRealigner in GATK, then removing duplicate reads with PICARD v1.64 MARKDUPLICATES (http://picard.sourceforge.net) and finally calling SNPs and indels using GATK's UnifiedGenotyper, which reports allele frequencies including SNPs and indels, with a minimum coverage cut-off for variant calls of 5 reads per variant allele. To check for indel mapping errors (e.g. especially near the ends of reads), bam files were viewed in TABLET 1.12.03.26 (Milne *et al.* 2010). Although sequencing results were high quality and several filtering steps were used, to address remaining sequencing or PCR error, only SNPs and indels of >1% frequency in at least one sampling location were included in subsequent analyses. Fixed SNPs were defined as having $\geq 99\%$ of reads matching an allele, and high-frequency SNPs and indels as $\geq 95\%$ or $\geq 85\%$ matching, respectively. The latter value was set somewhat lower because of increased uncertainty of calls due to possible mapping error. Lower-frequency SNPs and indels in the US were compared in two ways (i) by classification into functional categories and testing for overrepresentation or underrepresentation of 'hits' (number of polymorphic sites) and (ii) by analysing allele frequency changes since ground zero, grouping by distance, plant or functional category. Gene prediction was performed using PRODIGAL v2-50 (Hyatt *et al.* 2010) and by comparison with the fully annotated reference genome of *Ishikawaella* (GenBank Accession AP010872.1, AP010873.1) by viewing these in Artemis (Carver *et al.* 2012). Allele frequency change was defined as the change in % of reads with an alternate nucleotide (a SNP) compared with the % of reads with that nucleotide in the ground zero sample. Positive and negative values represent SNPs with increased or decreased frequency since ground zero, respectively, potentially reflecting genetic drift, positive selection or negative (purifying) selection. For analysis of functional categories (clusters of orthologous groups, COG; or gene ontology, GO), differences in allele frequency change were added and then normalized for the total length of genes in a category and for the number of samples. Basic statistical tests (likelihood ratio test, Spearman's rank correlation, Wilcoxon signed rank, etc.) were performed in Excel. Tests for multiple 'hits' to genes and GO categories were evaluated using permutation tests in GOWINDA v1.12 (Kofler & Schlötterer 2012) which controls for gene length differences, overlapping genes and performs multiple testing (false discovery rate, FDR) correction. Gowinda was run in snp mode with 100000 simulations using 10000 random SNPs to generate the null distribution, tested against GO categories downloaded from GOMINER v291 (Zeeberg *et al.* 2003). These methods act as an alternative to other neutrality tests that are not easily applied to short read data with low-level polymorphism within samples, for which haplotypes would be necessary. Partial genes (short ORFs) predicted in Prodigal and BLAST were examined in Tablet and Artemis for possible pseudogene differences compared with the reference genome for every GATK-predicted indel and SNP. Additive allele frequency changes, gene names and COG categories were displayed graphically using Circos (Krzywinski *et al.* 2009). Population genetics parameter estimation was performed as described in Data S1, Supporting information.

Results

Few functional differences between US and Japanese Ishikawaella strains

The US *Ishikawaella* genomes from all invasive *M. cribraria* were identical in gene order and orientation to the published Japanese *Ishikawaella* genome from the soybean pest insect *M. punctatissima* (Nikoh *et al.* 2011). The consensus US *Ishikawaella* genome was 46 bp longer than the Japanese reference genome due to six short indel differences. The US and Japanese *Ishikawaella* genomes differed by just 37 fixed and 10 high-frequency polymorphisms, including six indels and 41 substitutions. About half of these were in coding regions (14 nonsynonymous and 10 synonymous), and the remaining were in intergenic regions or pseudogenes (see Table S3, Supporting information). The 14 nonsynonymous differences affect genes that fall into three broad COG categories: six in cellular processes (O, M, P), four in information storage (J, K, L) and three in metabolism (C, H). Only two substitutions occurred in genes with putative symbiont nutritional roles (COG category H, coenzyme transport and metabolism): *cysG* (fused siroheme synthase 1,3-dimethyluroporphyrinogen) and *ribC* (riboflavin synthase subunit alpha). The *cysG* substitution was a synonymous C to A change at position 152423. The *ribC* substitution was a nonsynonymous T to G change at position 508418

producing a radical lysine to asparagine amino acid replacement. This substitution is in the lumazine-binding domain of the protein (by homology with *E. coli*, Liao *et al.* 2001) and so could affect the function of the enzyme. Careful examination of all intragenic SNP and indel regions (using BLAST searches) revealed no differential pseudogenization between the US and published Japanese pest genome. Plasmid sequences matched that of the published 9140 bp *Ishikawaella* plasmid from *M. punctatissima* (GenBank accession AP010873.1), except for a single nucleotide difference (A to T at position 2456, a synonymous change in methionine aminopeptidase).

Many polymorphisms, but no fixed differences across the US

No fixed allele differences were observed amongst US *Ishikawaella* samples. There were 164 low-frequency polymorphic sites with up to 27.5% allele frequency (i.e. percent of reads with alternate allele compared with reference allele from published Japanese *Ishikawaella*; Table S4, Supporting information). All US sampling locations showed some polymorphism (2–83 polymorphic alleles). All sites were diallelic (having only two alternate nucleotides). Of the 164 polymorphic sites, 95 occurred in only one sample and 69 were shared between two or more sampling locations. Forty-nine sites were polymorphic in the ground zero sample, and another 115 sites displayed new polymorphisms not present at ground zero. Ten were nonsense (stop) mutations (1.3–3.8% frequency), and five were frameshifts caused by 1 bp deletions (2.2–3.6% frequency). Amongst new polymorphic alleles, the ratio of nonsynonymous to synonymous sites was 65/20 (3.25:1).

Polymorphic sites are unevenly distributed amongst COG categories and genes

Most polymorphic alleles occurred in coding rather than intergenic regions (134 vs. 30), but most of the *Ishikawaella* genome is coding, such that allele distribution measured as alleles per nucleotide showed an even distribution between coding and intergenic regions. It was difficult to classify intergenic polymorphic sites because regulatory regions are poorly characterized in obligate nutritional symbionts. However, amongst alleles within coding regions, the distribution of SNPs was uneven between different COG categories after normalizing for the length of each category in base pairs (two-tailed likelihood ratio test for goodness-of-fit to null model of even distribution across COGs P-value = 0.0049, 16 d.f.). Lipid transport and metabolism (I), cell wall, membrane, envelope biogenesis (M) and signal transduction mechanisms (T) appear overrepresented, whereas replication, recombination and repair (L), carbohydrate transport and metabolism (G) and nucleotide transport and metabolism (F) appear underrepresented (Fig. S2, Supporting information). Despite the presence of 28 polymorphic sites in COG categories for amino acid and coenzyme metabolism (E and H), which are potential symbiont nutritional role-related genes, these categories displayed average numbers of polymorphic sites after controlling for the length of each COG.

The distribution of allele frequency change since ground zero was also uneven between different COG categories (Fig. 3). Negative allele frequency change was most dramatic for lipid transport and metabolism (I), cell cycle control, cell division and chromosome partitioning (D), intergenic regions (X) and amino acid metabolism (E). However, negative allele frequency change was not significantly different for the two plants (kudzu vs. soybeans two-tailed Wilcoxon signed rank test $Wcrit_{\ 0.05,9} = 35 > W = 27$, P-value = 0.652; Fig. 3). In contrast, positive allele frequency change was greatest for replication, recombination and repair (L), cell wall/membrane/envelope biogenesis (M), lipid transport and metabolism (I) and coenzyme transport and metabolism (H) (Fig. 3). In contrast to the negative allele frequency changes, positive allele frequency change was significantly different on kudzu vs. soybeans (two-tailed Wilcoxon signed rank test, $z_{crit} = -2.47$, $W = -105$, P-value = 0.0135).

Genes with two or more polymorphic sites (hereafter 'multiple hits') occurred with high frequency: 43% of polymorphisms (70 of 164 sites), representing 30 of all 84 genes with polymorphic sites. Of these 30 genes with multiple hits, 23 were new mutations not observed in the ground zero sample. Eight genes had more than two hits: *thrA*, *sdhB*, *accC*, *valS*, *clpB* each had 3; *sucD* had four; and *mreB* and *ribD* each had five. Many multiple hits occurred close together (57% were within 10 bp; median nc distance within genes 3.5 bp apart, mean = 184 bp; Table S5, Supporting information). Permutation tests in Gowinda supported overrepresentation in genes *ribD* and *mreB* (P-value = 0.00077), *sucD* (P-value = 0.0068) and *accC* (P-value = 0.01; Table S6, Supporting information) and several other enzymatic/protein complexes, processes or pathways including riboflavin biosynthesis (P-value = 0.0092; Table S7, Supporting information). Most essential amino acid biosynthesis pathways had multiple hits (e.g. Arg, Thr, Met, Ile, Val, Leu, Lys, Phe, Trp had 2, 3, 3, 2, 2, 4, 3, 2, 5 hits, respectively), but overrepresentation was not significant for these (Table S8, Supporting information).

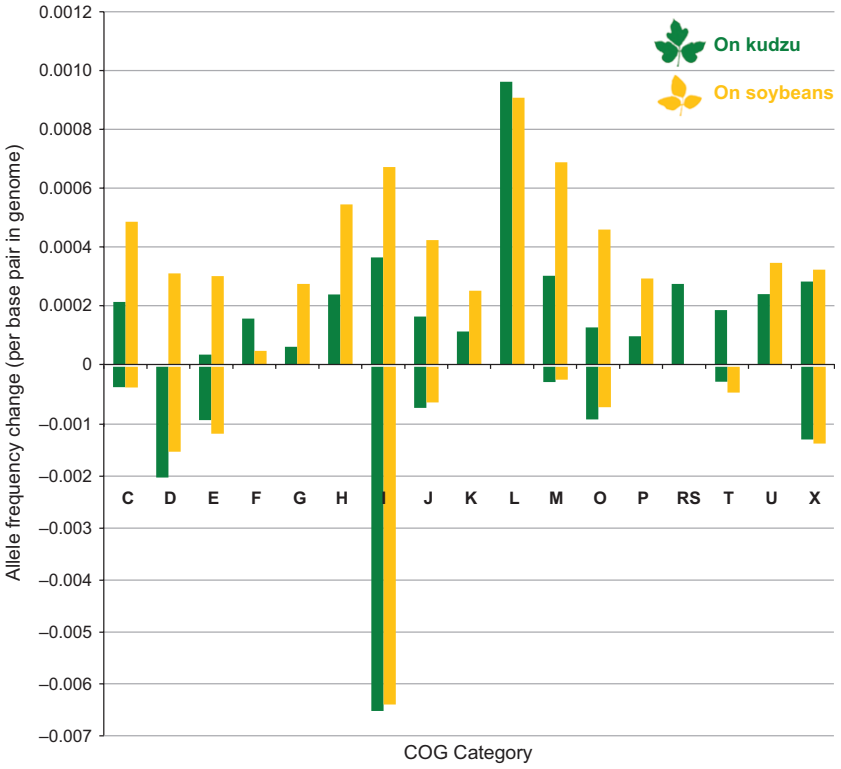

Fig. 3 Polymorphic variation in *Ishikawaella capsulata* sorted by clusters of orthologous groups category, showing uneven distribution of allele frequency change from ground zero depending on the plant (kudzu in green and soybeans in yellow). *Ishikawaella* samples show significantly more positive change on soybeans than on kudzu (see text), but no difference in negative change between the two plants. Negative *y*-axis is condensed for easier visualization. C = Energy production and conversion; D = Cell cycle control, cell division, chromosome partitioning; E = Amino acid transport and metabolism; F = Nucleotide transport and metabolism; G = Carbohydrate transport and metabolism; H = Coenzyme transport and metabolism; I = Lipid transport and metabolism; J = Translation, ribosomal structure and biogenesis; K = Transcription; L = Replication, recombination and repair; M = Cell wall/membrane/envelope biogenesis; O = Post-translational modification, protein turnover, chaperones; P = Inorganic ion transport and metabolism; RS = general/unknown; T = Signal transduction mechanisms; U = Intracellular trafficking, secretion and vesicular transport; X = Intergenic regions.

Allele frequency change varies with distance, host plant and COG category

Localities differed in overall (absolute) allele frequency change since ground zero and in the relative levels of nonsynonymous and synonymous change (see pie graphs Fig. 2), although the overall trend with respect to distance was complex (see Fig. S3, Supporting information), and a detailed analysis will depend on further sampling. On the whole, our results reflect an overall trend of increased allele frequency change in samples near the boundary of the 2010–2011 ranges. For example, excess of polymorphism (Tajima's D) vs. distance, as a measure of deviation from neutrality, was always negative (−2.6 and −4.4) with greatest variance in samples in mid-distances (Fig. S3, Supporting information). Negative values of Tajima's D are consistent with an expanding population or purifying selection.

The correlation of positive allele frequency change with distance also depended on plant: soybeans showed a more significant correlation with distance than did kudzu (one-way ANCOVA *P*-value 2.5E-17; both plants negatively correlated with distance: soybeans $\rho = -0.413$, *P*-value = 2.7E-7, 142 d.f.; kudzu $\rho = -0.185$, *P*-value = 0.028, 140 d.f.). These results reflect the pattern shown in Fig. 2, depicting three localities with the greatest allele frequency change also having the largest ratio of nonsynonymous to synonymous mutations, two of these being on soybeans. These patterns are consistent with what would be expected from positive selection or relaxed purifying selection.

Amino acid and cofactor biosynthesis genes (COG categories E and H) showed significantly higher allele frequency change for nonsynonymous sites on soybeans compared with kudzu (soybean = 45.3%, kudzu = 7.2%, one-tailed, two-sample unequal variance *t*-test *P*-value = 0.016). The net change for these COG categories was opposite: allele frequencies in genes

related to amino acid pathway alleles decreased, whereas coenzyme pathway alleles increased in frequency.

Alleles showing the greatest frequency change: frameshifts and recB

Polymorphic alleles were dispersed widely across the genome and occurred in diverse functional categories (Fig. 4). The majority of alleles (75%) showed low-frequency net positive change (Fig. 4 and see Table S9, Supporting information) including eight nonsense (stop) mutations. Adding change across all samples, only 10 alleles had >10% positive frequency change, whereas 41 alleles had >10% negative frequency change. The greatest negative allele changes include all five frameshifts and one nonsense mutation; these results are consistent with purifying selection acting on deleterious mutations. Three of the frameshift alleles present as ground zero polymorphisms were not observed in any other samples. Several of the most negative and most positive alleles were in intergenic regions (Fig. 4 and Table S9, Supporting information). Three intergenic alleles with large frequency change (two negative and one positive) were approximately 70–80 bp upstream of the 16S gene (positions 318557–318573), and the two others with large negative change were approximately 90–110 bp upstream of genes *ackA* and *dnaK* (positions 56145 and 357520).

The most positively changed allele was a nonsynonymous A to C SNP at position 166518 in *recB*. RecB, or Exonuclease V beta chain is part of the RecBCD complex involved in recombination and repair. The SNP we observed causes a conservative leucine to valine substitution at the *C*-terminal domain of the helicase portion of RecB, homologous to residue 735 in *E. coli* which lies between a 3_{10}-helix and a conserved beta strand (PDB ID:1W36 Singleton *et al.* 2004). We could find no reference to mutational experiments at this residue; however, mutations in the nearby conserved α–helix (motif VI) were shown to inhibit RecBCD helicase activity (see Dillingham & Kowalczykowski 2008; and references therein). Thus, it remains unclear whether this SNP could affect the enzyme function. Because RecBCD is involved in recombination and repair (Yu *et al.* 1998), we looked for a positive correlation between the number of new SNPs in a sample and the allele frequency of this *recB* SNP. The *recB* allele frequency was strongly positively correlated with new SNPs for soybeans, but not kudzu (Spearman's rank correlation: soybeans $\rho = 0.748$, P-value = 0.021, 7 d.f.; kudzu $\rho = 0.088$, P-value = 0.79, 10 d.f.; see Fig. S4, Supporting information). When considering all plants together, the frequency of the *recB* allele showed an increase with the number of new SNPs per sample and, to a lesser extent, with distance from ground zero (slope of linear regression 0.179 and 0.0184, respectively), whereas new SNPs did not increase as much with distance (slope of linear regression 0.00744). However, these correlations were not significant (linear regression for *recB* vs. SNPs $r^2 = 0.128$ P-value = 0.094, *recB* vs. distance $r^2 = 0.0821$ P-value = 0.19, and SNPs vs. distance $r^2 = 0.00333$ P-value = 0.79, all d.f. = 21). Principal component analysis supported this result (*recB* vs. SNPs and *recB* vs. distance: PC1 + PC2 = 0.81 of variance; SNPs vs. distance: PC3 = 0.19).

Several *ribD* (riboflavin deaminase reductase) mutations were high in positive allele frequency change. The highest of these occurred in five samples, showing >16% change added across samples. A nonsynonymous A to T mutation at position 290108 causes a radical leucine to phenylalanine substitution within a predicted conserved catalytic motif (Zn-binding site) and proton-donor site of the gene (by homology with RibD residues 51–57 in *E. coli*, Stenmark *et al.* 2007). Samples from soybeans showed more positive change in this allele. The remaining 3 *ribD* mutations were also in the Zn-binding domain, and four of seven samples with these alleles were from soybeans.

Discussion

Based on the observations that the genotype of the bacterial symbiont *Ishikawaella* determines the pest status of its host insect in Japan (Hosokawa *et al.* 2007) and that US host insects resemble nonpest insects in Asia (Eger *et al.* 2010; Jenkins & Eaton 2011), we searched for genetic changes that may be associated with emergence of the pest phenotype (i.e. the emergence on soybeans) in the recent US *M. cribraria* invasion. For *Ishikawaella* from ground zero and 23 other US locations on three plant species, we used deep sequencing and comparative genomics to test hypotheses on the symbiont's role in pest emergence (Fig. 1B). We also looked at changes in polymorphic allele frequency, expecting that in the absence of selection, new alleles, and changes in pre-existing alleles should be randomly distributed with respect to specific genes or COG categories, whereas alleles under selection may be clustered in the genome and change nonrandomly. If selection occurs, it may act between symbionts within a host or between symbiont populations within different host individuals, the latter potentially revealing genes that play a role in differential host survival. Below, we discuss our results in light of these predictions.

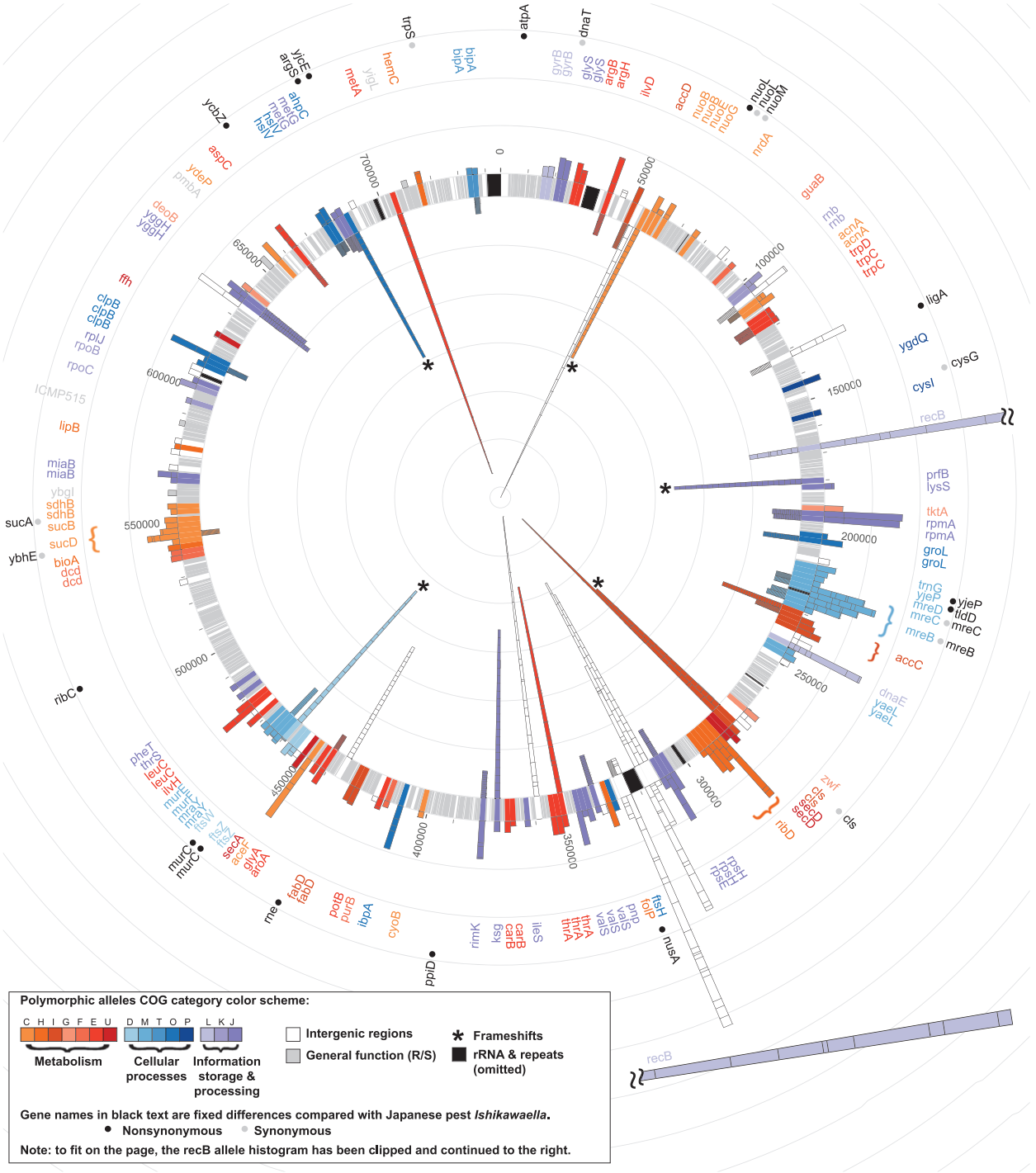

Fig. 4 Polymorphic allele frequency change since ground zero in US *Ishikawaella capsulata* mapped onto the genome, showing changes at individual loci in general functional categories. Histograms directed outward and inward represent positive and negative frequency change, respectively. Each stacked histogram box represents the allele frequency change compared with ground zero in a single sample location in the US. Histogram colours and coloured gene names represent different functional categories (see legend inset box and full clusters of orthologous groups category names in Fig. 3 legend). Gene names in black represent fixed substitutions or indels differing between the US consensus and published Japanese pest-conferring *Ishikawaella* (for simplicity of figure, fixed intergenic differences are not shown). Grey, white and black bars in the inner ring show for the remainder of the genome the coding, intergenic and rRNA/repeat regions, respectively. Asterisks indicate frameshift mutations. Thin grey concentric circles indicate histogram units in% allele frequency change compared with ground zero in steps of 10% for inward and 20% for outward histograms. Positions in the genome (0–745636) are indicated in grey numerals with tick marks just outside inner ring.

US insects harbour a strain of Ishikawaella that strongly resembles the Japanese pest-conferring symbiont

Our analyses demonstrated that overall nucleotide similarity and the predicted functional profile of the US symbiont *Ishikawaella* closely resembles the published pest-conferring *Ishikawaella*. We therefore hypothesize that the US *Ishikawaella* is functionally similar in its symbiont role to the Japanese pest version of *Ishikawaella*, with the exception of a single difference in the riboflavin synthase (*ribC*) gene. The significance of this single amino acid substitution in *ribC* is unclear. If it affects the efficiency of riboflavin biosynthesis, this might be important if riboflavin availability differs across legume species. While amino acid content in phloem can affect performance in the aphid-*Buchnera* symbiosis (Sandström & Pettersson 1994), to our knowledge, no studies examine content of riboflavin or vitamin/cofactors in phloem across legume species and their affect on a phloem feeder. It is not clear whether riboflavin could be limiting in this system. Nevertheless, this substitution is not polymorphic in the US, so presumably this change does not account for differences in plant use across US sampling locations or in the emergence of the pest phenotype in the first year after arrival.

To fully evaluate how US *Ishikawaella* confers the pest status, future symbiont swap experiments and genome sequencing should include other isolates from Asia for which the plant host range and pest status have been demonstrated experimentally. At present, these are not available. However, the genome from an *Ishikawaella* isolate from a *Megacopta* specimen from Fukuoka, Japan (AMVB data unpublished), is similar to that reported here from the US, with no further pseudogenization. The plant host range and pest status of this Fukuoka specimen could not be reliably established, although its location would suggest it may also be able to act as a soybean pest (Hosokawa *et al.* 2007; and references therein). We speculate, based on a molecular clock for *Buchnera* (see Moran *et al.* 2009; based on 19 substitutions per 270 years), that these three *Ishikawaella* genotypes may have diverged approximately 700 years ago, during a time of expansion of Asian soybean cultivation (Hymowitz & Kaizuma 1981). Taken together, our results complement recent experimental host plant range data (see Zhang *et al.* 2012) by providing a baseline genomic signature for the soybean-utilizing *Ishikawaella*.

How did the pest Ishikawaella originate?

While conclusive support for the Asian origins of the US *Megacopta-Ishikawaella* invasion will depend on more extensive sampling in Asia, the present finding that the US symbiont genome functionally resembles the pest-conferring symbiont from *M. punctatissima* in Japan, while the US insect resembles the Asian 'non-pest' *M. cribraria* (Eger *et al.* 2010; Jenkins & Eaton 2011) implicates some unusual event prior to or during the nascent invasion. Three possible nonexclusive scenarios are illustrated in Fig. 1B: (i) a symbiont swap, (ii) hybridization or (iii) novel mutations. The first scenario, a symbiont swap, may be likely here due to the extracellular condition in *Ishikawaella,* whereas it would be improbable in primary endosymbionts that are transmitted transovarially (Moran *et al.* 2008). Phylogenetic analyses for stinkbug symbionts show that symbiont swaps have been rare (Kikuchi *et al.* 2009). Nevertheless, the ease with which a swap was achieved in laboratory (Hosokawa *et al.* 2007) suggests it is possible. In nature, a swap may occur anywhere pest- and nonpest *Megacopta* species co-occur. Normally, newly hatched nymphs use their piercing and sucking mouthparts to probe for symbiont-containing capsules laid down with the eggs by their mothers (Fukatsu & Hosokawa 2002). However, if nymphs are disturbed or capsules are damaged or not found, nymphs rapidly disperse great distances (several metres) in search of other capsules (Hosokawa *et al.* 2008). Thus, if leaves bearing capsules from a different *Megacopta* species were found nearby, nymphs could easily encounter these and become inoculated with a different symbiont.

In considering the other scenarios spelled out in Fig. 1B, our data cannot rule out a hybridization event between the pest and nonpest *Megacopta* species or recombination between *Ishikawaella* strains. Hosokawa *et al.* (2007) report that F_1 hybrid insects from *M. cribraria* × *M. punctatissima* are fertile and produce viable F_2 offspring. Furthermore, Nikoh *et al.* (2011) show that *Ishikawaella* possess intact genes for recombination, which has been recently highlighted as an important and underestimated force in bacterial symbionts (Sachs *et al.* 2011 and references therein). If insects are sympatric in some localities, nonpests could hybridize with pests, or nymphs could probe multiple symbiont capsules becoming inoculated with both nonpest and pest *Ishikawaella* strains, which could then undergo recombination, perhaps by transformation.

An alternate explanation for pest emergence is novel 'pest' mutations in US *Ishikawaella* (Fig. 1B). This possibility would have important consequences for pest control in major soybean-growing regions (e.g. Argentina and Brazil) not yet infested with *Megacopta*. If novel pest mutations were shown to arise from nonpest *Ishikawaella* in the US, we might expect them to be able to evolve elsewhere; hence, controlling nonpest *Megacopta* spp. would be important. The pattern of

mutation we observed did not support this. Given the lack of fixed differences amongst US isolates (discussed below), it appears that the arriving *Ishikawaella* may already have the pest-conferring ability based on the functional similarity to the *Ishikawaella* that confers the pest status in Japan, as discussed earlier. The only substitution between the pest-conferring *Ishikawella* from *M. punctatissima* in Japan and the US *Ishikawaella* with a potential role in nutrient provisioning was a single *ribC* amino acid substitution that was already present at ground zero. Its presence in the arriving insect-*Ishikawaella* population does not support an origin of new pest mutations after arrival, although it is possible that it is related to the success of this pest. It is noteworthy that other independent mutations appear to have occurred after arrival within the same symbiont nutritional pathway (i.e. riboflavin biosynthesis), suggesting selection may be occurring on this trait. Further sampling in the US is needed to confirm or refute this hypothesis.

Genetic uniformity in US symbionts contrasts with plant use and morphological plasticity in insects

No fixed differences accumulated in the *Ishikawaella* isolates sampled from November 2009 at ground zero through October 2011 on soybeans, kudzu and *Lespedeza* spp. over an area of 120 000 km^2. The most important implication of this result is that it suggests the emergence on soybeans during the first year in the US was not due to a change in the symbiont, but was, perhaps, due simply to a lag between arrival and the colonization of soybeans by the pest-capable *M. cribraria*. Another implication of the lack of fixed differences in *Ishikawaella* is that in the US, the observed variation in host-plant distribution, colour, size of insects and appearance of the symbiotic-crypt bearing midgut (Fig. S1, Supporting information) are not related to symbiont genotype, but rather to either phenotypic plasticity or genetic variation in the insect. Results showing mitochondrial marker uniformity in *M. cribraria* across the US (2336 bp of mitochondrial DNA from 83 individuals from 36 counties) suggest the insects came from a single maternal lineage and may therefore have limited genetic diversity (Jenkins & Eaton 2011). The present data also produced *M. cribraria* mitochondrial data showing low polymorphism, no fixed differences (data not shown) and no obvious association between polymorphic alleles and phenotype (Table S1, Supporting information). Nevertheless, differences in nuclear genotype of the insect could still play a role. Indeed, while the present US *Megacopta-Ishikawaella* invaders are prolific soybean pests that only reproduce successfully on kudzu and soybeans (Zhang *et al.* 2012), there appears to be a range of success on different legume hosts in Asia (see Eger *et al.* 2010, and references therein). It is not clear to what extent the symbiont is responsible. More extensive genetic sampling of Asian variants and host-range experiments may help address the question of the phenotype–genotype relationships for both symbionts and insects in the Asian population. Experimental swaps of symbiont-containing capsules in the US *M. cribraria* with capsules from non-pests in Asia should demonstrate more definitively whether this *M. cribraria* strain is truly a functional 'non-pest' in the absence of its pest-conferring symbiont. This could confirm causality of the pest phenotype for the pest-like *Ishikawaella* genotype in the US.

While these data emphasize the potential role of the symbiont, there are other nonadaptive processes that may be involved. The most likely of these may be enemy release and host plant (phloem) nutritional profile differences either across the US or compared with Asian host plants. Indeed, enemy release (e.g. improved fitness in new habitat due to absence of parasites, predators, etc.) is a possibility here. In Asia, *Megacopta* species have evolved with numerous natural enemies not present in the US (Eger *et al.* 2010). Differential success in the US may also be affected by different nutrient profiles in the various US kudzu strains (Pappert *et al.* 2000) and soybean varieties. Alternatively, insect phenotypic plasticity may play a role. This could involve the symbiont, as suggested by MacDonald *et al.* (2011) who showed the nutritional phenotype of the pea aphid may be governed by the interaction of host metabolism and transporter genes that regulate nutrient supply to the *Buchnera* primary symbiont. Finally, we must consider whether a secondary symbiont could be involved. We found no evidence for a second primary nutritional symbiont in our sequence data from the gut, confirming the results of Hosokawa *et al.* (2007). The only other microbe present in our data was *Wolbachia* at very low sequence coverage (data not shown). Our data suggest that the *Wolbachia* present in US *Megacopta* are phylogenetically similar to strains in Supergroup A (Kikuchi & Fukatsu 2003), many of which cause cytoplasmic incompatibility. However, recent evidence suggests some members of this clade may be mutualists, affecting insect success on different plants (see Feldhaar 2011 and references therein). Further study of the *Wolbachia* found in US *Megacopta* will be presented elsewhere.

Deep sequencing reveals polymorphisms that may be under selection

While our data could only examine the early phase of this symbiotic invasion, our deep sequencing coverage

of *Ishikawaella* isolates was able to resolve small differences in low-frequency alleles. This strategy has been effective in finding rare HIV variants and somatic cancer cells (De Grassi *et al.* 2010) and should be quite important in studying invasions (Barrett & Schluter 2008). Here, we found patterns consistent with possible purifying and positive selection, but teasing apart the effects of drift and selection on these allele frequency changes is difficult with our current data. For example, some alleles that increased in frequency occurred as clusters of SNPs within genes or pathways. While our results are preliminary, we suggest some of these genes or pathways could be under positive selection, most notably genes involved in riboflavin biosynthesis (*ribC* and *ribD* with five polymorphic sites in the US, and one fixed difference between the Japanese and US genomes) and the *mreBCD* alleles (eight polymorphic sites, two fixed differences) involved in cell shape. The *ribD* allele frequencies showed significantly different change on kudzu vs. soybeans, suggesting possible differential importance of riboflavin (vitamin B_2) biosynthesis by *Ishikawaella* on these plants. While these early data are not yet corroborated by experimental studies, we suggest *ribD*—and by extension the entire riboflavin pathway—may be worth examining in the future as a candidate gene or pathway involved in success on soybeans in the US.

The *recB* allele, an A to C SNP at position 166518, displayed the greatest positive frequency change. Interestingly, we found a positive relationship between the *recB* allele frequency and the number new SNPs (most of which would presumably be deleterious). If this relationship was causative, we might expect the *recB* allele would be selectively disadvantageous and subject to purifying selection. However, studies show that moderately deleterious alleles may rise to high frequency due to demographically driven drift ('surfing') at the wave front of an expanding population (Travis *et al.* 2007; Excoffier & Ray 2008; François *et al.* 2010). This raises the question of whether the *recB* allele could be increasing in frequency even though it may be deleterious. Further study would be needed to address this possibility.

Conclusions

Previous studies have raised the question of when and how microbial symbionts might control their host's phenotype (Tsuchida *et al.* 2004; Gerardo & Wilson 2011; Simms & Porter 2012). We suggest this question is especially important for invasive species (Lee 2002) and likely for symbionts of Hemipteran insects. We examined a system that has become a serious recent invasive pest in the US (Zhang *et al.* 2012) and where the genotype of the symbiont may cause a pest phenotype in its host, as has been shown experimentally for related strains in Japan (Hosokawa *et al.* 2007). We were primarily interested in determining the role of the symbiont in causing the emergence of the pest phenotype in the US. We found strong evidence that the genotype of US *Ishikawaella* in *M. cribraria* functionally resembles the *Ishikawaella* strain that confers the pest status in Japan, in *M. punctatissima*. Thus, we suggest the initial population of invading insects was probably able to infest soybeans, a surprising result given the resemblance of the host insect to nonpests in Asia. We hypothesize that some combination of symbiont swap or previous insect hybridization may have occurred prior to the invasion potentially facilitating the US outbreak. We also discuss several other nonadaptive factors, such as enemy release, that may play a role. Furthermore, our results clearly show that the observed differences in phenotype and morphology across the US cannot be due to *Ishikawaella* genotype differences. These findings point to future experiments with the US insect that should help determine causality of symbiont facilitation in this system (e.g. use of Asian specimens for symbiont swaps and host plant range tests). The current results provide a genomic baseline for comparative functional genomics to accompany such experiments. Finally, our analyses of SNP frequency changes and distribution suggest potential pathways of interest, such as riboflavin biosynthesis, that may be important for this insect to thrive on soybeans.

Acknowledgements

We thank N. M. Gerardo, T. M. Jenkins, D. R. Suiter, J. K. Greene and M. L. Allen for providing insects and helping with collection. We thank J. T. Van Leuven for the use of Perl scripts. We thank N. M. Gerardo, T. Spribille, D. Vanderpool, the editor (J. Russell), and three anonymous reviewers for suggestions which improved the manuscript. This work was funded by a USDA AFRI grant (2011-67013-30090) to J.P.M.

References

Andersson DI (2008) Shrinking bacterial genomes. *Microbe*, **3**, 124–130.

Barrett RDH, Schluter D (2008) Adaptation from standing genetic variation. *Trends in Ecology and Evolution*, **23**, 38–44.

Baumann P (2005) Biology bacteriocyte-associated endosymbionts of plant sap-sucking insects. *Annual Reviews in Microbiology*, **59**, 155–189.

Buchner P (1965) *Endosymbiosis of Animals with Plant Microorganisms*, Interscience Publishers, New York, New York.

Carver T, Harris SR, Berriman M, Parkhill J, McQuillan JA (2012) Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics*, **28**, 464–469.

Clay K, Holah J (1999) Fungal endophyte symbiosis and plant diversity in successional fields. *Science*, **285**, 1742–1744.

Dale C, Moran NA (2006) Molecular interactions between bacterial symbionts and their hosts. *Cell*, **126**, 453–465.

De Grassi A, Segala C, Iannelli F *et al.* (2010) Ultradeep sequencing of a human ultraconserved region reveals somatic and constitutional genomic instability. *PLoS Biology*, **8**, e1000275.

DePristo MA, Banks E, Poplin R *et al.* (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, **43**, 491–498.

Desprez-Loustau M-L, Robin C, Buée M *et al.* (2007) The fungal dimension of biological invasions. *Trends in Ecology and Evolution*, **22**, 472–480.

Dillingham MS, Kowalczykowski SC (2008) RecBCD enzyme and the repair of double-stranded DNA breaks. *Microbiology and Molecular Biology Reviews*, **72**, 642–671.

Eger Jr JE, Ames LM, Suiter DR, Jenkins TM, Rider DA, Halbert SE (2010) Occurrence of the Old World bug *Megacopta cribraria* (Fabricius) (Heteroptera: Plataspidae) in Georgia: a serious home invader and potential legume pest. *Insecta Mundi*, **0121**, 1–11.

Excoffier L, Ray N (2008) Surfing during population expansions promotes genetic revolutions and structuration. *Trends in Ecology and Evolution*, **23**, 347–351.

Feldhaar H (2011) Bacterial symbionts as mediators of ecologically important traits of insect hosts. *Ecological Entomology*, **36**, 533–543.

François O, Currat M, Ray N, Han E, Excoffier L, Novembre J (2010) Principal component analysis under population genetic models of range expansion and admixture. *Molecular Biology and Evolution*, **27**, 1257–1268.

Fukatsu T, Hosokawa T (2002) Capsule-transmitted gut symbiotic bacterium of the Japanese common plataspid stinkbug, *Megacopta punctatissima*. *Applied and Environmental Microbiology*, **68**, 389–396.

Gerardo NM, Wilson ACC (2011) The power of paired genomes. *Molecular Ecology*, **20**, 2038–2040.

Hosokawa T, Kikuchi Y, Meng XY, Fukatsu T (2005) The making of symbiont capsule in the plataspid stinkbug *Megacopta punctatissima*. *FEMS Microbiology Ecology*, **54**, 471–477.

Hosokawa T, Kikuchi Y, Nikoh N, Shimada M, Fukatsu T (2006) Strict host-symbiont cospeciation and reductive genome evolution in insect gut bacteria. *PLoS Biology*, **4**, e337.

Hosokawa T, Kikuchi Y, Shimada M, Fukatsu T (2007) Obligate symbiont involved in pest status of host insect. *Proceedings of the Royal Society of London. Series B*, **274**, 1979–1984.

Hosokawa T, Kikuchi Y, Shimada M, Fukatsu T (2008) Symbiont acquisition alters behaviour of stinkbug nymphs. *Biology Letters*, **4**, 45–48.

Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*, **11**, 119.

Hymowitz T, Kaizuma N (1981) Soybean seed protein electrophoresis profiles from 15 Asian countries or regions: hypotheses on paths of dissemination of soybeans from China. *Economic Botany*, **35**, 10–23.

Janson EM, Stireman III JO, Singer MS, Abbot P (2008) Phytophagous insect-microbe mutualisms and adaptive evolutionary diversification. *Evolution*, **62**, 997–1012.

Jenkins TMJ, Eaton TD (2011) Population genetic baseline of the first plataspid stink bug symbiosis (Hemiptera: Heteroptera: Plataspidae) reported in North America. *Insects*, **2**, 264–272.

Kikuchi Y, Fukatsu T (2003) Diversity of *Wolbachia* endosymbionts in Heteropteran bugs. *Applied and Environmental Microbiology*, **69**, 6082–6090.

Kikuchi A, Kobayashi H (2010) Effect of injury by adult *Megacopta punctatissima* (Montandon) (Hemiptera: Plataspidae) on the growth of soybean during the vegetative stage of growth. *Japanese Journal of Applied Entomology and Zoology*, **54**, 37–43.

Kikuchi Y, Hosokawa T, Nikoh N, Meng X-Y, Kamagata Y, Fukatsu T (2009) Host-symbiont co-speciation and reductive genome evolution in gut symbiotic bacteria of acanthosomatid stinkbugs. *BMC Biology*, **7**, 2.

Kofler R, Schlötterer C (2012) Gowinda: unbiased analysis of gene set enrichment for genome-wide association studies. *Bioinformatics*, **28**, 2084–2085.

Krzywinski M, Schein J, Birol İ *et al.* (2009) Circos: an information aesthetic for comparative genomics. *Genome Research*, **19**, 1639–1645.

Lee CE (2002) Evolutionary genetics of invasive species. *Trends in Ecology and Evolution*, **17**, 386–391.

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics*, **25**, 1754–1760.

Li H, Handsaker B, Wysoker A *et al.* (2009) The Sequence alignment/map (SAM) format and SAMtools. *Bioinformatics*, **25**, 2078–2079.

Liao DI, Wawrzak Z, Calabrese JC, Viitanen PV, Jordan DB (2001) Crystal structure of riboflavin synthase. *Structure*, **9**, 399–408.

Lichman E (2010) Invisible invaders: non-pathogenic invasive microbes in aquatic and terrestrial ecosystems. *Ecology Letters*, **13**, 1560–1572.

MacDonald SJ, Thomas GH, Douglas AE (2011) Genetic and metabolic determinants of nutritional phenotype in an insect-bacterial symbiosis. *Molecular Ecology*, **20**, 2073–2084.

McCutcheon JP, Moran NA (2010) Functional convergence in reduced genomes of bacterial symbionts spanning 200 My of evolution. *Genome Biology and Evolution*, **2**, 708–718.

McCutcheon JP, Moran NA (2011) Extreme genome reduction in symbiotic bacteria. *Nature Reviews Microbiology*, **10**, 13–26.

McKenna A, Hanna M, Banks E *et al.* (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, **20**, 1297–1303.

Milne I, Bayer M, Cardle L *et al.* (2010) Tablet—next generation sequence assembly visualization. *Bioinformatics*, **26**, 401–402.

Mira A, Ochman H, Moran NA (2001) Deletional bias and the evolution of bacterial genomes. *Trends in Genetics*, **17**, 589–596.

Moran NA, McCutcheon JP, Nakabachi A (2008) Genomics and evolution of heritable bacterial symbionts. *Annual Review of Genetics*, **42**, 165–190.

Moran NA, McLaughlin HJ, Sorek R (2009) The dynamics and time scale of ongoing genomic erosion in symbiotic bacteria. *Science*, **323**, 379–382.

Nikoh N, Hosokawa T, Oshima K, Hattori M, Fukatsu T (2011) Reductive evolution of bacterial genome in insect gut environment. *Genome Biology and Evolution*, **3**, 702–714.

Pappert RA, Hamrick JL, Donovan LA (2000) Genetic variation in *Pueraria lobata* (Fabaceae), an introduced, clonal, invasive plant of the southeastern United States. *American Journal of Botany*, **87**, 1240–1245.

Pimentel D, Zuniga R, Morrison D (2005) Update on the environmental and economic costs associated with alien-invasive species in the United States. *Ecological Economics*, **52**, 273–288.

Sachs JL, Essenberg CJ, Turcotte MM (2011) New paradigms for the evolution of beneficial infections. *Trends in Ecology and Evolution*, **26**, 202–209.

Sandström J, Pettersson J (1994) Amino acid composition of phloem sap and the relation to intraspecific variation in pea aphid (*Acyrthosiphon pisum*) performance. *Journal of Insect Physiology*, **40**, 947–955.

Shah MA, Reshi ZA, Khasa DP (2009) *Arbuscular mycorrhizas*: drivers or passengers of alien plant invasion. *Botanical Review*, **75**, 397–417.

Simms EL, Porter SS (2012) Transcriptomic insights into mechanisms of symbiotic cooperation. *Molecular Ecology*, **21**, 4665–4668.

Singleton MR, Dillingham MS, Gaudier M, Kowalczykowski SC, Wigley DB (2004) Crystal structure of RecBCD enzyme reveals a machine for processing DNA breaks. *Nature*, **432**, 187–193.

Son MS, Taylor RK (2011) Preparing DNA libraries for multiplexed paired-end deep sequencing for Illumina GA sequencers. *Current Protocols in Microbiology*, Chapter 1, Unit 1E.4.

Stenmark P, Moche M, Gurmu D, Nordlund P (2007) The crystal structure of the bifunctional deaminase/reductase RibD of the riboflavin biosynthetic pathway in *Escherichia coli*: implications for the reductive mechanism. *Journal of Molecular Biology*, **373**, 48–64.

Suiter DR, Eger Jr JE, Gardner WA *et al.* (2010) Discovery and distribution of *Megacopta cribraria* (Hemiptera: Heteroptera: Plataspidae) in northeast Georgia. *Journal of Integrated Pest Management*, **1**, 1–4.

Travis JM, Münkemüller T, Burton OJ, Best A, Dytham C, Johst K (2007) Deleterious mutations can surf to high densities on the wave front of an expanding population. *Molecular Biology and Evolution*, **24**, 2334–2343.

Tsuchida T, Koga R, Fukatsu T (2004) Host plant specialization governed by facultative symbiont. *Science*, **303**, 1989.

Vogel KJ, Moran NA (2011) Sources of variation in dietary requirements in an obligate nutritional symbiosis. *Proceedings of the Royal Society of London Series B*, **278**, 115–121.

Yu M, Souaya J, Julin DA (1998) The 30-kDa *C*-terminal domain of the RecB protein is critical for the nuclease activity, but not the helicase activity, of the RecBCD enzyme from *Escherichia coli*. *Proceedings of the National Academy of Sciences USA*, **95**, 981–986.

Zeeberg BR, Feng W, Wang G *et al.* (2003) GoMiner: a resource for biological interpretation of genomic and proteomic data. *Genome Biology*, **4**, R28.

Zhang Y, Hanula JL, Horn S (2012) The biology and preliminary host range of *Megacopta cribraria* (Heteroptera: Plataspidae) and its impact on kudzu growth. *Environmental Entomology*, **41**, 40–50.

A.M.V.B. and J.P.M. designed the research and wrote the paper, L.Y.H. and A.M.V.B. collected specimens and did field research, K.G.N. performed dissections and qPCR, C.M.B. designed primers and conducted PCR, A.M.V.B. performed the research and analyzed the data.

Data accessibility

Sample information: Table S1. Indexed US *Ishikawaella* reads (.fastq files) and final genome sequence assembly with mapped reads (.bam files): Sequence Read Archive database (NCBI) accession numbers SRR784403, SRR784402, SRR745157, SRR710525. Raw SNP data: Dryad doi:10.5061/dryad.266 h9.

Supporting information

Additional supporting information may be found in the online version of this article.

Data S1 Detailed methods describing genomic library preparation, demultiplexing, filtering, assembly and population genetic parameter estimation.

Table S1 Details on *Megacopta* sample collection including host plant, date, location and insect colour.

Table S2 Illumina read data for barcoded samples, showing numbers of raw and filtered reads, percent of reads mapped to *Ishikawaella* and coverage for each barcoded index.

Table S3 Fixed substitutions and indels (and high frequency polymorphisms) between US *Ishikawaella* and the Japanese pest-conferring *Ishikawaella* from *Megacopta punctatissima*.

Table S4 Polymorphic alleles in US *Ishikawaella* with position, gene names, clusters of orthologous groups categories (or 'x' for intergenic) and allele frequency.

Table S5 Polymorphic alleles in US *Ishikawaella* that occur as 'multiple hits' (two or more per gene) along with distance in base pairs between neighbouring SNPs/indels.

Table S6 Statistical tests for overrepresentation of genes for polymorphic alleles in US *Ishikawaella*, calculated using GOWINDA v1.12 permutation tests.

Table S7 Statistical tests for overrepresentation of gene ontology (GO) categories for polymorphic alleles in US *Ishikawaella* calculated using GOWINDA.

Table S8 Statistical tests for overrepresentation of pathways for polymorphic alleles in US *Ishikawaella* calculated using GOWINDA.

Table S9 Polymorphic allele frequency change relative to ground zero in US *Ishikawaella* for each sample location and summed across locations.

Fig. S1 *Megacopta cribraria* morphological variation (external body and gut) amongst US samples and several Asian specimens.

Fig. S2 Histogram showing number of polymorphic SNPs/indels in US *Ishikawaella* per clusters of orthologous groups (COG) category (normalized for length of COG in genome).

Fig. S3 Further details on polymorphic allele frequency change correlation with distance and Pie and Tajima's D results.

Fig. S4 Scatter plots showing the relationship between the recB allele frequency change, number of new SNPs and distance from ground zero for all sample locations with and without separation by host plant.