Supplementary Material File S1 for article in Biology Letters:

Intrinsic pre-zygotic reproductive isolation of distantly related pea aphid host races

Varvara Fazalova[1*], Bruno Nevado[2], Ailsa McLean[1] and H. Charles J. Godfray[1]

*Corresponding author

[1]Department of Zoology, University of Oxford,

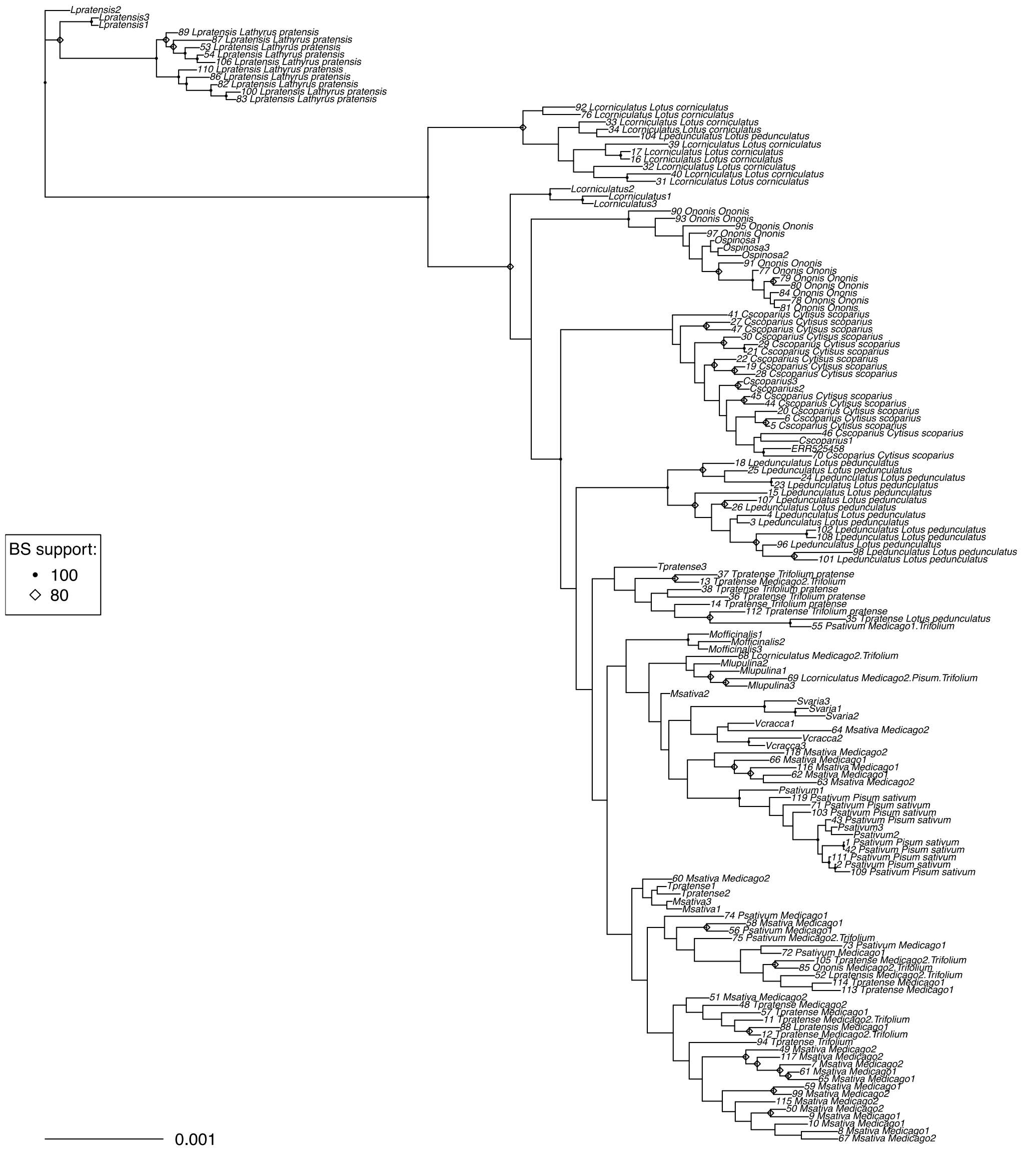
[2]Department of Plant Sciences, University of Oxford

File S1. Phylogenetic reconstruction (for pea aphid host races).

Genetic clustering of individuals from the same pea aphid host races from distant locations has shown using microsatellite markers [1–3]. However, phylogenetic reconstruction based on few markers can be misleading due to incomplete lineage sorting or ongoing gene flow. Thus, in order to choose distantly and closely related biotypes, we constructed a phylogeny based on available genome-wide data (Duvaux et al. 2015 and the SpeciAphid project). The dataset of Duvaux et al. [4] contained targeted resequencing of ~500 chemosensory genes from 120 individuals representing eight host races collected in the UK, while the data from the SpeciAphid project consisted of whole-genome resequencing data (unpublished, used with permission of authors, AphidBase http://bipaa.genouest.org/is/aphidbase/) from 33 individuals sampled from 11 host-plant races collected in France.

We downloaded the raw Illumina data from GenBank (Bioprojects PRJEB6325 and PRJNA255937), trimmed low-quality ends and adaptors using cutadapt v 1.8.3 [5], and mapped reads to the pea aphid genome version Acyr_2.0 using bwa mem v. 0.7.12 [6] with default parameter values. Mapped data in bam format were realigned around indels using the GATK v. 3.4 [7], and used to perform per-individual SNP calling in samtools v. 1.2 [8] with a minimum quality threshold of 20 for both mapping and base-quality. Obtained SNPs were filtered to remove low quality SNP calls (< 15); SNPs with low depth (< 8 reads); SNPs near indels (< 3 bp); and heterozygous SNPs with less than 2 reads supporting each allele. We used the option to report homozygous-reference blocks for each individual with a minimum depth of 8 reads (bcftools call –g8), and the resulting vcf files were converted into fasta format using custom scripts. We concatenated the resulting fasta files for each scaffold after removing sites with more than 80% missing data, in effect selecting only the sequences for the ca. 500 chemosensory genes used in Duvaux et al. (2015). The concatenated dataset (570,278 bp) was used for Maximum-Likelihood phylogenetic inference with raxml v 8.0 [9] using the GTR model with Gamma-distributed rate variation. We performed 100 rapid bootstrap runs to gauge node support for the best-scoring maximum likelihood (ML) tree obtained. The phylogenetic tree obtained confirmed that French and British host races belong to the same genetic groups, likely reflecting the importance of long-distant dispersal rather than in-situ diversification in determining phylogeographic patterns within this species complex (**Figure S1.1**).

Figure S1.1 Compete maximum likelihood tree for pea aphid host races. Pea aphid phylogeny based on ca. 500 chemosensory genes for clones collected in the UK and France. Clones from France have format “Plant_host Sample_number”, the UK clones: “Clone_Number Collection_Plant Host-race_assignment”. Host-race assignment (can be more than to one host race). is according to supplementary files Duvaux_CNV-PeaAphid_Sup.Tables from [4].


Lpratensis2
Lpratensis3
Lpratensis1
89 Lpratensis Lathyrus pratensis
87 Lpratensis Lathyrus pratensis
53 Lpratensis Lathyrus pratensis
54 Lpratensis Lathyrus pratensis
106 Lpratensis Lathyrus pratensis
110 Lpratensis Lathyrus pratensis
86 Lpratensis Lathyrus pratensis
82 Lpratensis Lathyrus pratensis
100 Lpratensis Lathyrus pratensis
83 Lpratensis Lathyrus pratensis
92 Lcorniculatus Lotus corniculatus
76 Lcorniculatus Lotus corniculatus
33 Lcorniculatus Lotus corniculatus
34 Lcorniculatus Lotus corniculatus
104 Lpedunculatus Lotus pedunculatus
39 Lcorniculatus Lotus corniculatus
17 Lcorniculatus Lotus corniculatus
16 Lcorniculatus Lotus corniculatus
32 Lcorniculatus Lotus corniculatus
40 Lcorniculatus Lotus corniculatus
31 Lcorniculatus Lotus corniculatus
Lcorniculatus2
Lcorniculatus1
Lcorniculatus3
90 Ononis Ononis
93 Ononis Ononis
95 Ononis Ononis
97 Ononis Ononis
Ospinosa1
Ospinosa3
Ospinosa2
91 Ononis Ononis
77 Ononis Ononis
79 Ononis Ononis
80 Ononis Ononis
84 Ononis Ononis
78 Ononis Ononis
81 Ononis Ononis
41 Cscoparius Cytisus scoparius
27 Cscoparius Cytisus scoparius
47 Cscoparius Cytisus scoparius
30 Cscoparius Cytisus scoparius
29 Cscoparius Cytisus scoparius
21 Cscoparius Cytisus scoparius
22 Cscoparius Cytisus scoparius
19 Cscoparius Cytisus scoparius
28 Cscoparius Cytisus scoparius
Cscoparius3
Cscoparius2
45 Cscoparius Cytisus scoparius
44 Cscoparius Cytisus scoparius
20 Cscoparius Cytisus scoparius
6 Cscoparius Cytisus scoparius
5 Cscoparius Cytisus scoparius
46 Cscoparius Cytisus scoparius
Cscoparius1
ERR525458
70 Cscoparius Cytisus scoparius
18 Lpedunculatus Lotus pedunculatus
25 Lpedunculatus Lotus pedunculatus
24 Lpedunculatus Lotus pedunculatus
23 Lpedunculatus Lotus pedunculatus
15 Lpedunculatus Lotus pedunculatus
107 Lpedunculatus Lotus pedunculatus
26 Lpedunculatus Lotus pedunculatus
4 Lpedunculatus Lotus pedunculatus
3 Lpedunculatus Lotus pedunculatus
102 Lpedunculatus Lotus pedunculatus
108 Lpedunculatus Lotus pedunculatus
96 Lpedunculatus Lotus pedunculatus
98 Lpedunculatus Lotus pedunculatus
101 Lpedunculatus Lotus pedunculatus
Tpratense3
37 Tpratense Trifolium pratense
13 Tpratense Medicago2.Trifolium
38 Tpratense Trifolium pratense
36 Tpratense Trifolium pratense
14 Tpratense Trifolium pratense
112 Tpratense Trifolium pratense
35 Tpratense Lotus pedunculatus
55 Psativum Medicago1.Trifolium
Mofficinalis1
Mofficinalis2
Mofficinalis3
68 Lcorniculatus Medicago2.Trifolium
Mlupulina2
Mlupulina1
69 Lcorniculatus Medicago2.Pisum.Trifolium
Mlupulina3
Msativa2
Svaria3
Svaria1
Svaria2
Vcracca1
64 Msativa Medicago2
Vcracca2
Vcracca3
118 Msativa Medicago2
66 Msativa Medicago1
116 Msativa Medicago1
62 Msativa Medicago1
63 Msativa Medicago2
Psativum1
119 Psativum Pisum sativum
71 Psativum Pisum sativum
103 Psativum Pisum sativum
43 Psativum Pisum sativum
Psativum3
Psativum2
1 Psativum Pisum sativum
42 Psativum Pisum sativum
111 Psativum Pisum sativum
2 Psativum Pisum sativum
109 Psativum Pisum sativum
60 Msativa Medicago2
Tpratense1
Tpratense2
Msativa3
Msativa1
74 Psativum Medicago1
58 Msativa Medicago1
56 Psativum Medicago1
75 Psativum Medicago2.Trifolium
73 Psativum Medicago1
72 Psativum Medicago1
105 Tpratense Medicago2.Trifolium
85 Ononis Medicago2.Trifolium
52 Lpratensis Medicago2.Trifolium
114 Tpratense Medicago1
113 Tpratense Medicago1
51 Msativa Medicago2
48 Tpratense Medicago2
57 Tpratense Medicago1
11 Tpratense Medicago2.Trifolium
88 Lpratensis Medicago1
12 Tpratense Medicago2.Trifolium
94 Tpratense Trifolium
49 Msativa Medicago2
117 Msativa Medicago2
7 Msativa Medicago2
61 Msativa Medicago1
65 Msativa Medicago1
59 Msativa Medicago1
99 Msativa Medicago2
115 Msativa Medicago2
50 Msativa Medicago2
9 Msativa Medicago1
10 Msativa Medicago1
8 Msativa Medicago1
67 Msativa Medicago2
BS support:
• 100
◇ 80
0.001

References

1. Peccoud J, Ollivier A, Plantegenest M, Simon J-C. 2009 A continuum of genetic divergence from sympatric host races to species in the pea aphid complex. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 7495–500. (doi:10.1073/pnas.0811117106)
2. Ferrari J, West JA, Via S, Godfray HCJ, Al ET. 2012 Population genetic structure and secondary symbionts in host-associated populations of the pea aphid complex. *Evolution (N. Y).* **66**, 375–390. (doi:10.5061/dryad.8qb00)

3. Henry LM, Peccoud J, Simon J-C, Hadfield JD, Maiden MJC, Ferrari J, Godfray HCJ. 2013 Horizontally transmitted symbionts and host colonization of ecological niches. *Curr. Biol.* **23**, 1713–7. (doi:10.1016/j.cub.2013.07.029)
4. Duvaux L, Geissmann Q, Gharbi K, Zhou J-J, Ferrari J, Smadja CM, Butlin RK. 2015 Dynamics of Copy Number Variation in Host Races of the Pea Aphid. *Mol. Biol. Evol.* **32**, 63–80. (doi:10.1093/molbev/msu266)
5. Martin M. 2011 Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10. (doi:10.14806/ej.17.1.200)
6. Li H, Durbin R. 2009 Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760.
7. McKenna A *et al.* 2010 The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–303. (doi:10.1101/gr.107524.110)
8. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009 The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079.
9. Stamatakis A. 2014 RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313. (doi:10.1093/bioinformatics/btu033)