Phylogenetics and quantitative genetics of host specificity in aphid parasitoids in the genus *Aphelinus*

Keith Hopper¹, Kristen Kuhn¹, Li Qiyun¹,², Randall Wisser³, Shawn Polson³, Sara Oppenheim⁴, James Woolley⁵, John Heraty⁶, Vladimir Gokhman⁷, Kim Hoelmer¹,¹¹, George Heimpel⁸, William O’Neill⁹, David Voetglin¹⁰, Kathryn Lanier¹, Joshua Rhoades¹, Robert Kondos¹, Dominique Coutinot¹¹, Guy Mercadier¹¹, Maria Roche¹¹, Nathalie Ramualde¹¹

¹USDA-ARS-Newark, Delaware; ²Jilin Academy of Agricultural Sciences, China; ³University of Delaware; ⁴American Museum of Natural History; ⁵Texas A&M University; ⁶University of California Riverside; ⁷Moscow State University, Russia; ⁸University of Minnesota; ⁹Ohio State University; ¹⁰Illinois Natural History Survey; ¹¹USDA-ARS-Montpellier, France

Funding: USDA-ARS, USDA-AFRI, NSF, North Central Soybean Research Program
Biology of the genus *Aphelinus*

- Parasitoids of aphids
- 90 described species
- Development time ~ 3 weeks at 20°C
- Synovigenic (produce eggs as adults)
- Host feed for nutrients for egg production
- Egg load 8-30; lifetime fecundity 100-200
- Long lived adults (2-4 weeks in lab)
- Small (~1mm), weak fliers
- Haplodiploid sex determination
Species and populations studied

**varipes complex**
- **Aphelinus kurdjumovi**
- **Aphelinus hordei**
- **Aphelinus varipes**
- **Aphelinus near varipes**
- **Aphelinus albipodus**
- **Aphelinus atriplicis**
- **Aphelinus certus**
- **Aphelinus near certus**

**asychis complex**
- **Aphelinus asychis**
- **Aphelinus sinensis**

**mali complex**
- **Aphelinus coreae**
- **Aphelinus glycinis**
- **Aphelinus rhamni**

* New to science:
Experiments on host specificity

Specificity measurement
- 100 aphids (mixed stages) single species
- Aphids on host plants in cages
- One female wasp per cage
- 24 hour exposure
- 20 C, 16:8 hours light:dark in plant growth chambers
- Measured number of aphids parasitized, adult progeny, sex ratio
- Host acceptance plus suitability

Behavioral basis of specificity
- 10 aphids on excised host plant leaf
- Small arena (100 mm x 40 mm)
- Individual female wasps
- 25 min of observations
- Measured number of aphids approached, stung, oviposited, and mummified, as well as behavioral defenses
Experiments on host specificity

Designs

- **Specificity measurement**
  - 17 aphid species in 8 genera
  - 11 host plant species in 9 families
  - 37 parasitoid populations in 15 species
  - 8-20 replicates per combination of aphid and wasp species
  - >3000 wasps measured

- **Behavioral basis of specificity**
  - 9-12 aphid species in several genera
  - 4 parasitoid species
  - 10-20 replicates
  - half of replicates aphids dissected, half aphids reared

- **Genetic basis of specificity with interspecific backcross**
  - Cross: *A. certus* ♂ x [*A. certus* ♀ x *A. atriplicis* ♂]
  - Host: *Diuraphis noxia*
  - 320 backcross progeny
  - measured specificity of parasitism
Variation in host specificity among *Aphelinus* species

Pipeline for genome and transcriptome sequencing, assembly and annotation

• DNA from single male (haploid) parasitoid (~100 ng)
  – Extracted with Qiagen DNEasy tissue kit
• Sequencing
  – Illumina HiSeq 2500, 100-150 nt paired-end
  – 15-40 Gbases per genome, 50-140x coverage
• Assembly
  – CLC Bio Genomics Workbench
  – Algorithm: de Bruijn graphs, word size 25, bubble size 50
  – Quality statistics: N\textsubscript{50}, N contigs, BUSCO-arthropod scores
• Annotation
  – Augustus for gene-finding, \textit{Nasonia vitripennis} models
  – Blastp of NCBI RefSeq to find homologs
  – Blast2go for gene ontologies
  – InterProScan for domain identifications
  – OrthoDB for ortholog/paralog clusters among \textit{Aphelinus} species
• RNA from 200-400 females and 200-400 males
  – Reads mapped to genome assembly with putative gene annotations
### Genome/transcriptome statistics for ten *Aphelinus* species

<table>
<thead>
<tr>
<th>Species complex:</th>
<th>asychis</th>
<th>atriplicis</th>
<th>certus</th>
<th>near certus</th>
<th>hordei</th>
<th>kurdjumovi</th>
<th>varipes</th>
<th>mali</th>
<th>sinensis</th>
<th>atriplicis</th>
<th>certus</th>
<th>near certus</th>
<th>hordei</th>
<th>kurdjumovi</th>
<th>varipes</th>
<th>coreae</th>
<th>glycinis</th>
<th>rhamni</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flow cytometry size (Mb)</td>
<td>-</td>
<td>353</td>
<td>362</td>
<td>344</td>
<td>393</td>
<td>346</td>
<td>333</td>
<td>439</td>
<td>432</td>
<td>483</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assembly size (Mb)</td>
<td>345</td>
<td>275</td>
<td>284</td>
<td>275</td>
<td>301</td>
<td>289</td>
<td>298</td>
<td>323</td>
<td>332</td>
<td>354</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coverage (reads per base)</td>
<td>71</td>
<td>53</td>
<td>103</td>
<td>-</td>
<td>121</td>
<td>141</td>
<td>72</td>
<td>134</td>
<td>118</td>
<td>123</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contig size - N50 (kb)</td>
<td>12</td>
<td>13</td>
<td>11</td>
<td>11</td>
<td>9</td>
<td>11</td>
<td>5</td>
<td>6</td>
<td>3</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contig size - maximum (kb)</td>
<td>172</td>
<td>134</td>
<td>118</td>
<td>120</td>
<td>117</td>
<td>135</td>
<td>95</td>
<td>130</td>
<td>83</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N contigs &gt;1000 nt</td>
<td>43,874</td>
<td>36,145</td>
<td>38,764</td>
<td>38,814</td>
<td>42,565</td>
<td>39,479</td>
<td>56,960</td>
<td>57,757</td>
<td>71,882</td>
<td>81,601</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BUSCO Arthropods (%)</td>
<td>92</td>
<td>93</td>
<td>92</td>
<td>-</td>
<td>92</td>
<td>92</td>
<td>91</td>
<td>91</td>
<td>89</td>
<td>87</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N genes</td>
<td>29,192</td>
<td>25,386</td>
<td>25,254</td>
<td>24,657</td>
<td>25,531</td>
<td>24,689</td>
<td>26,871</td>
<td>28,460</td>
<td>30,421</td>
<td>32,210</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N chemoreception proteins</td>
<td>96</td>
<td>94</td>
<td>93</td>
<td>91</td>
<td>84</td>
<td>88</td>
<td>82</td>
<td>86</td>
<td>93</td>
<td>85</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N cytochrome p450 proteins</td>
<td>116</td>
<td>109</td>
<td>103</td>
<td>105</td>
<td>104</td>
<td>103</td>
<td>116</td>
<td>104</td>
<td>131</td>
<td>121</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N venom proteins</td>
<td>115</td>
<td>112</td>
<td>111</td>
<td>107</td>
<td>113</td>
<td>109</td>
<td>121</td>
<td>134</td>
<td>134</td>
<td>143</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percent putative genes with:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>transcripts</td>
<td>-</td>
<td>92</td>
<td>93</td>
<td>-</td>
<td>99</td>
<td>100</td>
<td>97</td>
<td>95</td>
<td>96</td>
<td>107</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>blastp hits</td>
<td>73</td>
<td>80</td>
<td>78</td>
<td>80</td>
<td>78</td>
<td>79</td>
<td>77</td>
<td>74</td>
<td>76</td>
<td>74</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GO annotations</td>
<td>30</td>
<td>38</td>
<td>36</td>
<td>36</td>
<td>34</td>
<td>34</td>
<td>34</td>
<td>32</td>
<td>32</td>
<td>32</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Molecular phylogeny of *Aphelinus*

Parasitism mapped on *Aphelinus* molecular phylogeny

**Conserved pattern:** *Aphis glycines*

**Divergent pattern:** *Diuraphis noxia*

Parasitism level:

- 0 to 10
- 10 to 20
- 20 to 30
- 30 to 40
- 40 to 50
- 50 to 60
- 60 to 70
- 70 to 80
- 80 to 90
- 90 to 100
- 100

Crosses for QTL mapping

**Aphelinus certus**
- Female: A/T p G/C
- Male: A/T p G/C

**Aphelinus atriplicis**
- Male: G/C p A/T

**F₁ hybrid**
- Female: G/C p A/T
- Male: A/T p G/C

**Backcross hybrid**
- Female: G/C p A/T
- Male: A/T p G/C

- **non-recombinant**
  - Female: A/T p G/C
  - Male: A/T p A/T

- **recombinant**
  - Female: G/C p G/C
  - Male: G/C p A/T
Restriction-site Associated Sequence Polymorphisms

Reduced Representation bioinformatic pipeline

- 10$/sample for library preparation
- Up to 394 samples per Illumina HiSeq lane
- 100 nt single-end $1,350/lane ➔ $4/sample
Calling SNP versus haplotypes

- Even with a good reference (maize) single nucleotide polymorphisms can be miscalled
- Paralogs confounded

LOD profiles and QTL locations in *A. atriplicis* (R/qtl*)

Sizes of effects of QTL and interactions

Analysis of variance for QTL affecting parasitism of *Diuraphis noxia* by backcross females (*A. certus* ♂ x [*A. certus* ♀ x *A. atriplicis* ♂])

**ANOVA for model:**

<table>
<thead>
<tr>
<th></th>
<th>df</th>
<th>F</th>
<th>Pr(F)</th>
<th>% variance explained</th>
<th>LOD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>11</td>
<td>17.6</td>
<td>&lt;0.00001</td>
<td>39</td>
<td>33.9</td>
</tr>
<tr>
<td>Error</td>
<td>308</td>
<td>4.7</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**ANOVA for individual QTL and interactions:**

| Q1 | 1  | 12.7 | 0.0004  | 3  | 12.7 | 34  |
| Q2 | 2  | 25.8 | <0.00001| 10 | 25.8 | 9   |
| Q3 | 2  | 22.3 | <0.00001| 9  | 22.3 | 7   |
| Q4 | 2  | 19.2 | <0.00001| 8  | 19.2 | 6   |
| Q5 | 1  | 13.4 | 0.0003  | 3  | 13.4 | 105 |
| Q6 | 2  | 17.4 | <0.00001| 7  | 17.4 | 130 |
| Q7 | 2  | 13.2 | <0.00001| 5  | 13.2 | 11  |
| Q8 | 2  | 18.6 | <0.00001| 7  | 18.6 | 9   |
| Q2 x Q3 | 1 | 45.1 | <0.00001| 9  | 45.1 | -   |
| Q4 x Q6 | 1 | 30.3 | <0.00001| 6  | 30.3 | -   |
| Q7 x Q8 | 1 | 20.7 | <0.00001| 4  | 20.7 | -   |
Sequence-divergent genes between *A. atriplicis* and *A. certus* (CLC Bio, Blast2GO)

2% of genes

log number of genes

rate of non-synonymous versus synonymous substitutions (Ka/Ks)

- **total**: 491 genes
- **no blast hits**: 106 genes
- **with blast hits**: 52 genes
- **GO mapped**: 55 genes
- **GO annot.**: 278 genes

15 potential chemoreceptor proteins

40 potential chemoreceptor proteins
Annotation statistics for genes that are divergent in sequence or expression between *A. atriplicis* and *A. certus*

**Sequence divergent genes** (Ka/Ks ≥ 3) expressed only in females

- Total: 29
- With Blast (without hits): 11
- With Blast Hits: 3
- With Mapping: 6
- With GO Annotation: 9

**Expression divergent genes** (≥ 100x) in females

- Total: 21
- 2 potential chemoreceptor proteins: 4
- Several potential transcription factor proteins: 4
- Several potential transcription factor proteins: 11
Chromosomal Fluorescent In Situ Hybridization (FISH) to determine gene locations

- 28S ribosomal gene (red)
- DAPI stained chromosomes (blue)
- Structured illumination microscopy
Genotypic candidate genes with Kompetitive Allele Specific PCR

**KASP protocol (LGC, UK)**
- Two allele-specific forward primers
- One common reverse primer
- Mastermix:
  - Taq polymerase
  - Fluorescently labelled oligonucleotides
- Match with 5-prime tail of allele-specific primer
- PCR proceeds if correct sequence present in template DNA
- Appropriately labelled oligo releases fluor for detection during RT-PCR

**Primers for KASP**
- Sequenced genes in parents and grandparents
- Analyzed for SNPs between *A. atriplicis* and *A. certus*
Effect sizes and locations of divergent genes affecting parasitism of *D. noxia* (Blast, InterproScan, R/qtl)

<table>
<thead>
<tr>
<th>Gene description</th>
<th>Chromosome</th>
<th>Position (cM, LOD=6)</th>
<th>QTL</th>
<th>n parasitized aphids</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>serine protease precursor</td>
<td>1</td>
<td>34-63</td>
<td>Q1</td>
<td>0.4</td>
<td>1.70E-08</td>
</tr>
<tr>
<td>serine protease precursor</td>
<td>1</td>
<td>45-67</td>
<td>Q1</td>
<td>-</td>
<td>0.29</td>
</tr>
<tr>
<td>signal peptide, transmembrane*</td>
<td>1</td>
<td>52-111</td>
<td>Q1</td>
<td>0.7</td>
<td>0.002</td>
</tr>
<tr>
<td>serine protease</td>
<td>1</td>
<td>66-110</td>
<td>Q1</td>
<td>0.5</td>
<td>5.03E-06</td>
</tr>
<tr>
<td>signal peptide, transmembrane*</td>
<td>1</td>
<td>108-111</td>
<td>Q1</td>
<td>0.6</td>
<td>0.003</td>
</tr>
<tr>
<td>DNA pol. b2 domain-containing protein</td>
<td>1</td>
<td>109-111</td>
<td>Q1</td>
<td>0.7</td>
<td>0.002</td>
</tr>
<tr>
<td>protein yellow</td>
<td>1</td>
<td>145-186</td>
<td>Q2&amp;3</td>
<td>-</td>
<td>0.001</td>
</tr>
<tr>
<td>trypsin-1-like</td>
<td>1</td>
<td>164-167</td>
<td>Q3</td>
<td>-</td>
<td>0.01</td>
</tr>
<tr>
<td>serine protease precursor</td>
<td>1</td>
<td>220-261</td>
<td>Q4</td>
<td>0.2</td>
<td>0.05</td>
</tr>
<tr>
<td>-</td>
<td>2</td>
<td>54-95</td>
<td>Q5&amp;6</td>
<td>-</td>
<td>0.19</td>
</tr>
<tr>
<td>transmembrane*</td>
<td>2</td>
<td>64-106</td>
<td>Q5&amp;6</td>
<td>-</td>
<td>0.22</td>
</tr>
<tr>
<td>trypsin 1</td>
<td>2</td>
<td>178-209</td>
<td>Q6</td>
<td>-</td>
<td>0.39</td>
</tr>
<tr>
<td>hypothetical protein</td>
<td>3</td>
<td>34-71</td>
<td>Q7&amp;8</td>
<td>0.6</td>
<td>1.44E-06</td>
</tr>
<tr>
<td>-</td>
<td>4</td>
<td>0-30</td>
<td>none</td>
<td>0.6</td>
<td>0.0004</td>
</tr>
<tr>
<td>chymotrypsin-like protein</td>
<td>4</td>
<td>49-94</td>
<td>none</td>
<td>-</td>
<td>0.34</td>
</tr>
</tbody>
</table>

* inferred from domain analysis with InterProScan in Blast2GO
Tissue-specific expression of candidate genes

Scanning electron micrograph club of female antenna

Confocal image of sensilla labelled with red fluorescent probe for mRNA of candidate chemoreceptor gene
New NIFA-AFRI Grant

Objectives:

1. Use MAGIC RILs for high-resolution map of QTL affecting parasitism of novel host species
2. Test whether genes that diverge in sequence and/or expression fall under QTL
3. Test whether genes under QTL are expressed in predicted tissues
4. Compare transcriptomes of 15 Aphelinus species to test relationship between divergence in gene sequences and/or expression levels and host specificity

Job searches:

Postdoctoral Scientist ($54k/year)
Research Assistant ($44k/year)