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

Keith Hopper[1], Kristen Kuhn[1], Li Qiyun[1,2], Randall Wisser[3], Shawn Polson[3], Sara Oppenheim[4], James Woolley[5], John Heraty[6], Vladimir Gokhman[7], Kim Hoelmer[1,11], George Heimpel[8], William O'Neill[9], David Voetglin[10], Kathryn Lanier[1], Joshua Rhoades[1], Robert Kondos[1], Dominique Coutinot[11], Guy Mercadier[11], Maria Roche[11], Nathalie Ramualde[11]

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

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

200 µm

Species and populations studied

varipes complex

- 🟩 *Aphelinus kurdjumovi*
- 🟢 *Aphelinus hordei*
- ⬟ *Aphelinus varipes*
- ⬟ *Aphelinus* near *varipes*
- ◆ *Aphelinus albipodus*
- ★ *Aphelinus atriplicis*
- ▲ *Aphelinus certus*
- ▼ *Aphelinus* near *certus*

mali complex

- 🟨 *Aphelinus coreae**
- 🟡 *Aphelinus glycinis**
- 🔺 *Aphelinus rhamni**

asychis complex

- 🟧 *Aphelinus asychis*
- 🟠 *Aphelinus sinensis**
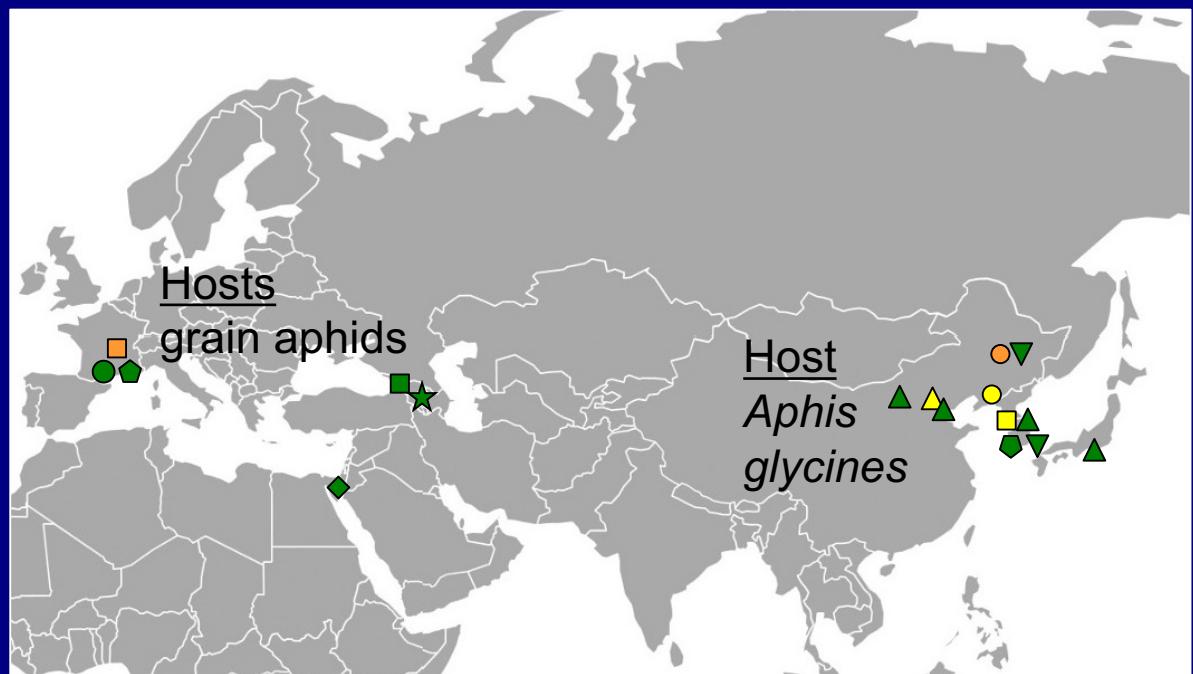
Hosts
grain aphids

Host
Aphis glycines

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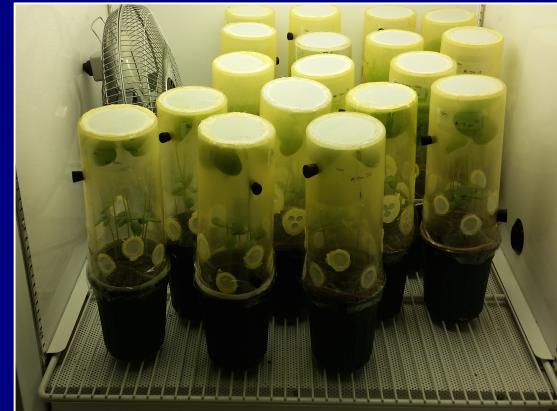
Behavioral basis of specificity
- – 10 aphids on excised host plant leaf
- – Small arena (100 mm x 40 mm)
- – Individual female wasps
- – 25 min of observations
- – Measured number of aphids
   approached, stung,
   oviposited, and mummified,
   as well as behavioral defenses

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
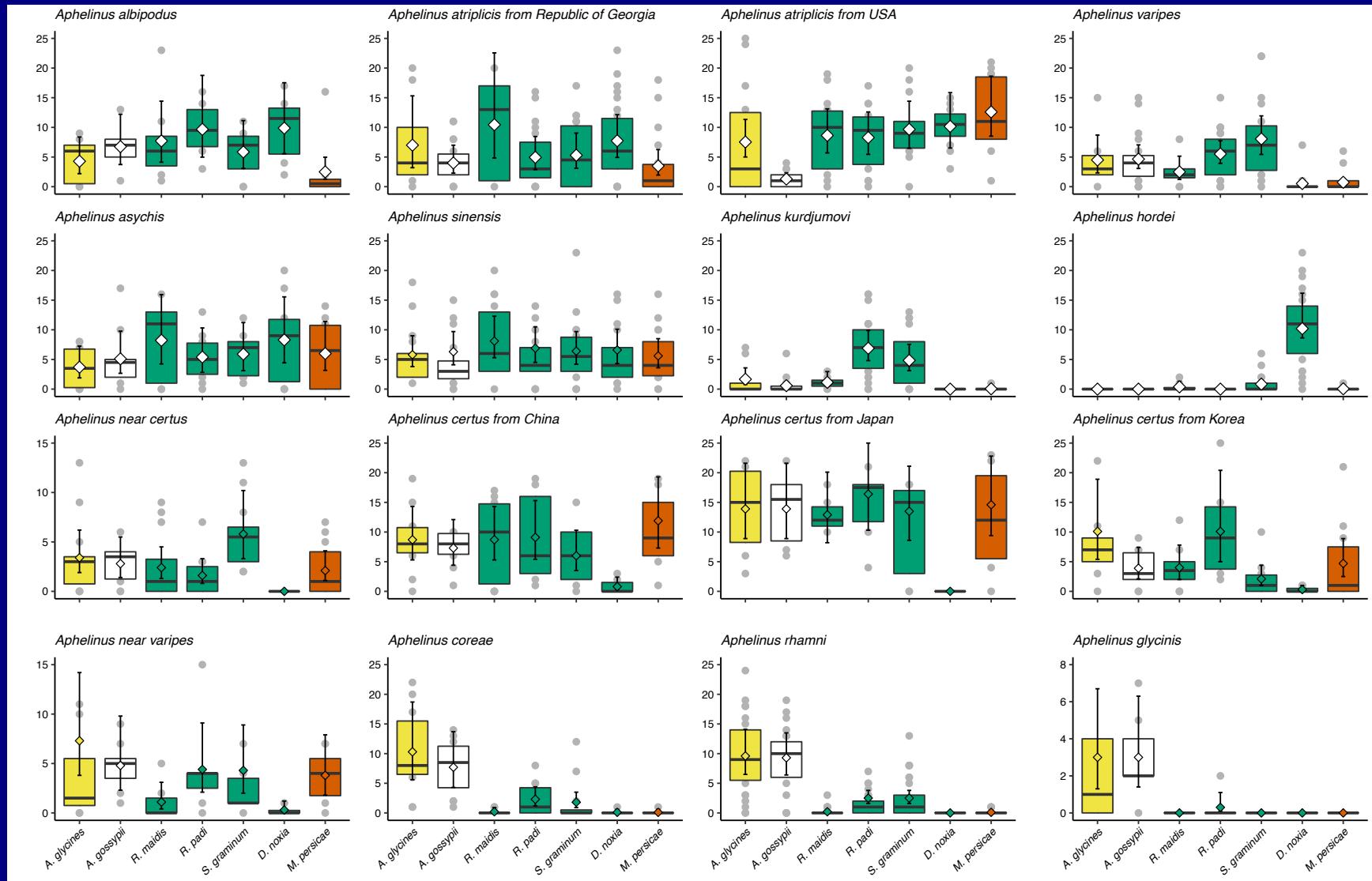
Hopper, K. R., K. Lanier, J. Rhoades, D. Coutinot, G. Mercadier, N. Ramualde, and M. Roche. 2017. Host specificity of *Aphelinus* species considered for introduction to control Diuraphis noxia. Biological Control: In press.

# 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_{50}, N contigs, BUSCO-arthropod scores**
- **Annotation**
  - **Augustus for gene-finding, *Nasonia vitripennis* models**
  - **Blastp of NCBI RefSeq to find homologs**
  - **Blast2go for gene ontologies**
  - **InterProScan for domain identifications**
  - **OrthoDB for ortholog/paralog clusters among *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

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

Molecular phylogeny of *Aphelinus*

Maximum likelihood analysis of sequence variation in amino acid sequences for 114 genes with RAxML (Stamatakis A. 2014. Bioinformatics 30:1312-1313). Species complexes: *asychis*, *varipes*, *mali*.
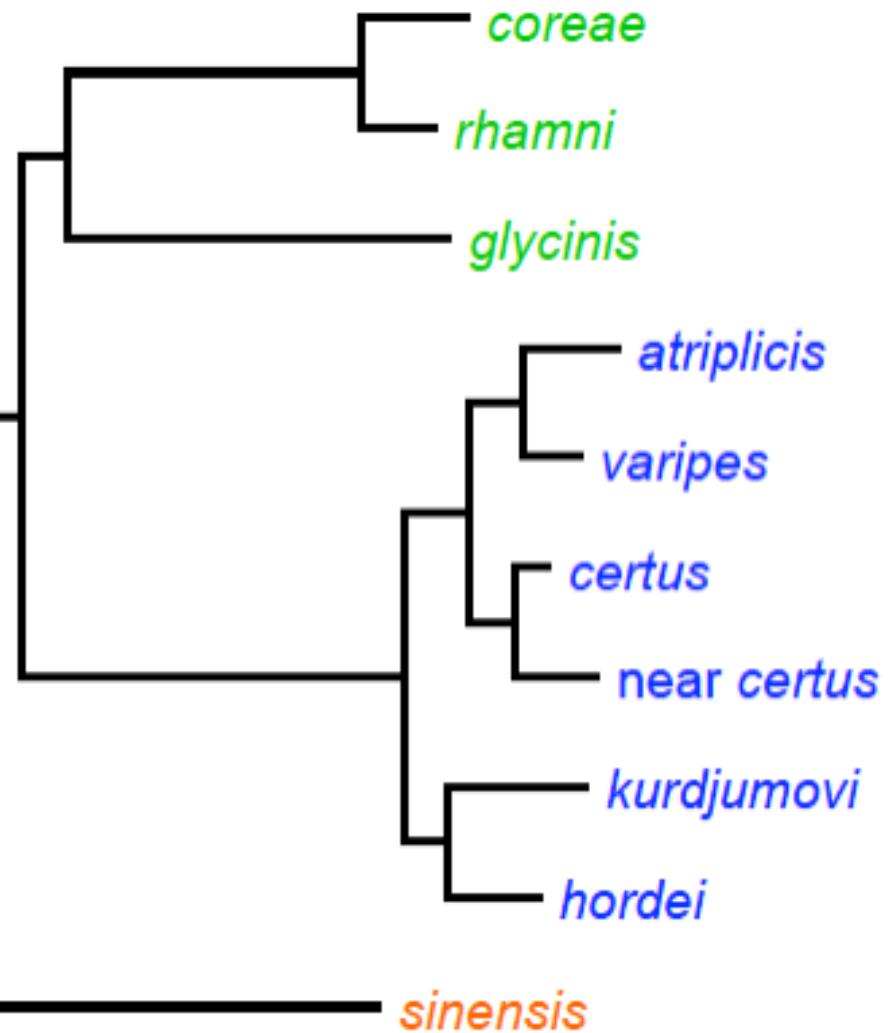
Parasitism mapped on *Aphelinus* molecular phylogeny

Conserved pattern:
Aphis glycines
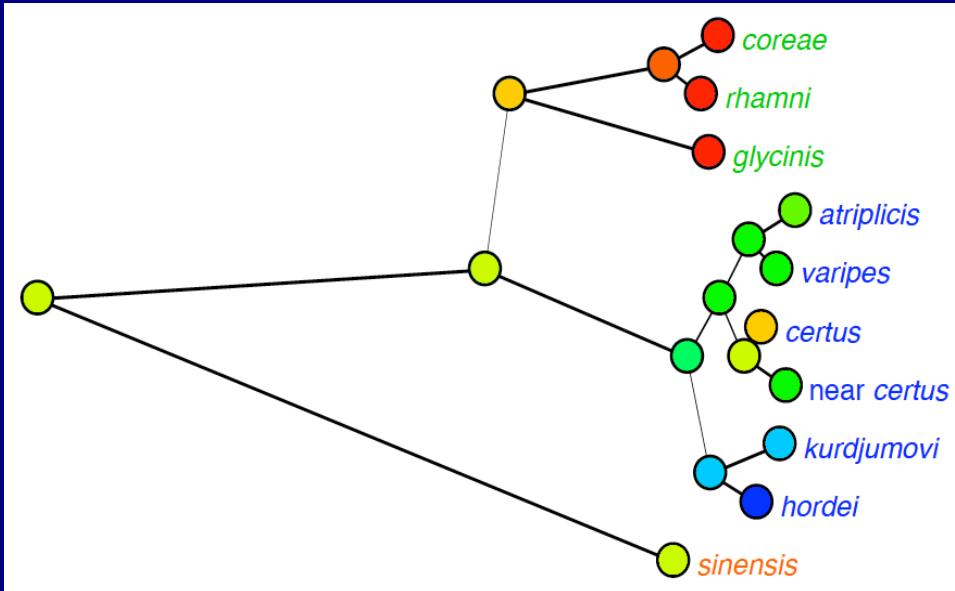
Divergent pattern:
Diuraphis noxia

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

Top phylogeny taxa: coreae, rhamni, glycinis, atriplicis, varipes, certus, near certus, kurdjumovi, hordei, sinensis

Bottom phylogeny taxa: coreae, rhamni, glycinis, atriplicis, varipes, certus, near certus, kurdjumovi, hordei, sinensis
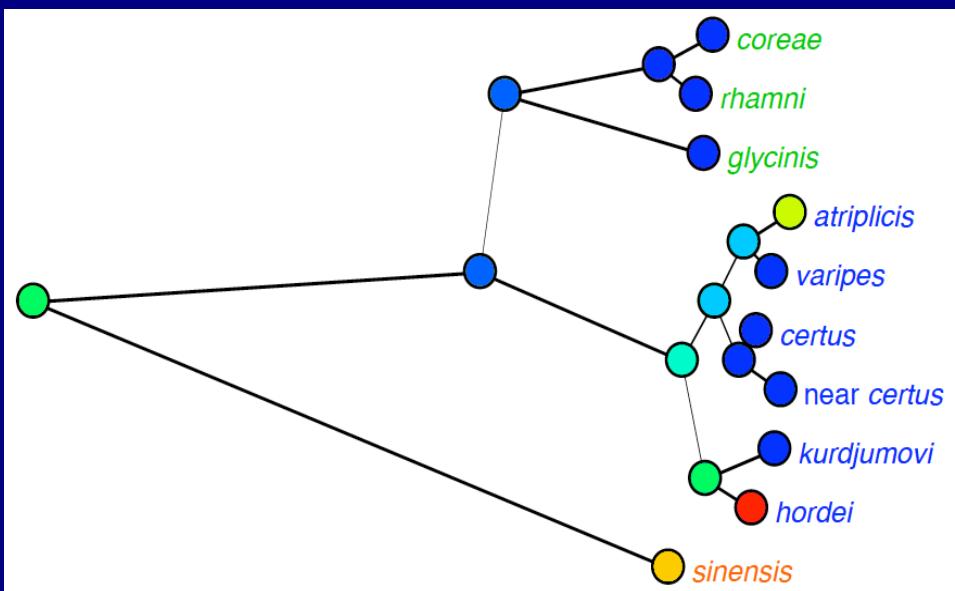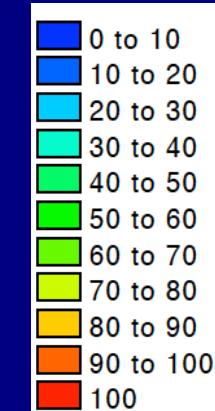
Parsimony reconstruction of ancestral states with Mesquite (Maddison WP & Maddison DR. 2007. Mesquite: a modular system for evolutionary analysis. Version 2.0 http://mesquiteproject.org.)
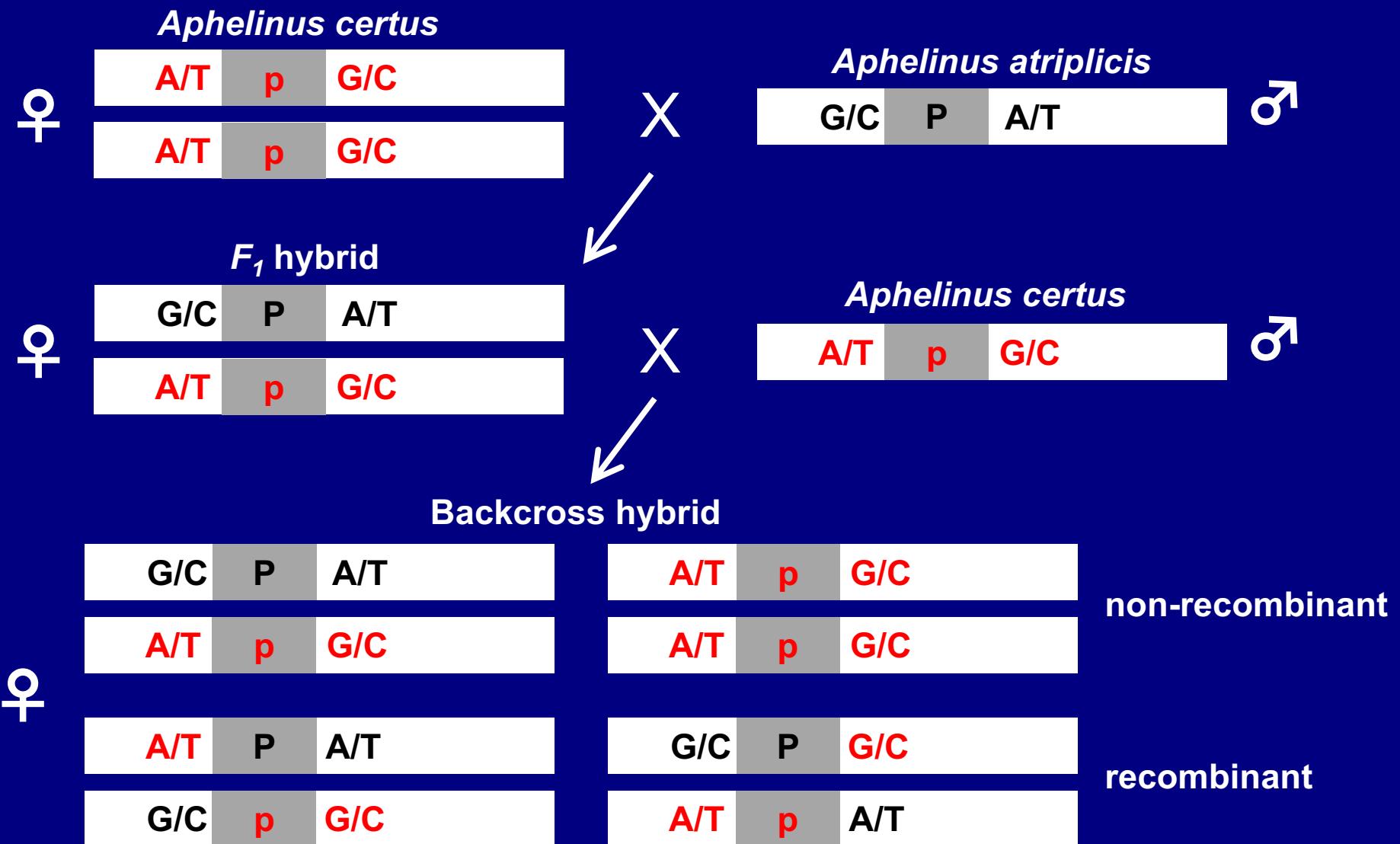
Restriction-site Associated Sequence Polymorphisms
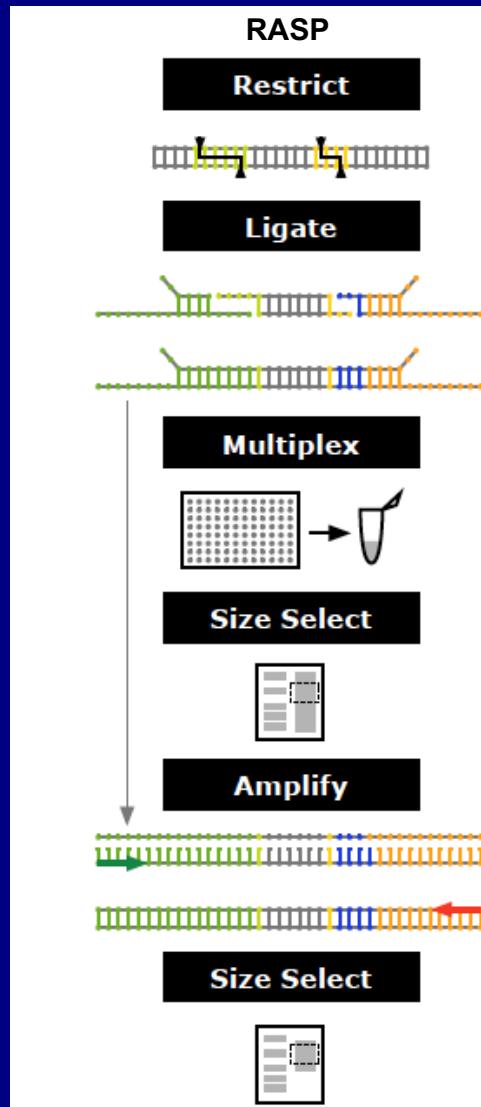
Reduced Representation bioinformatic pipeline
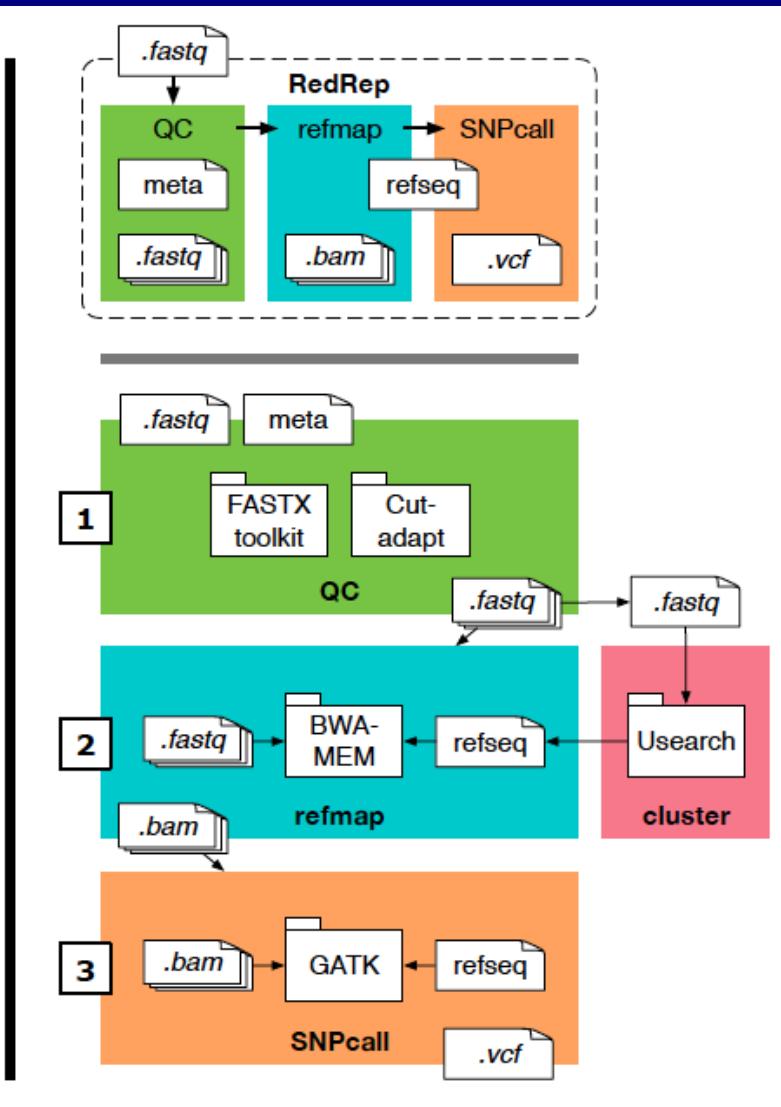

- 10$/sample for library preparation

- Up to 394 samples per Illumina HiSeq lane

- 100 nt single-end $1,350/lane →$4/sample

- RedRep code available at: https://github.com/UD-CBCB/RedRep

Calling SNP versus haplotypes

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

- **Paralogs confounded**

- **LocHap-GBS code available at:**

    http://www.compgenome.org/lochap/GBS/

Manching, H., S. Sengupta, K. R. Hopper, S. W. Polson, Y. Ji, and R. J. Wisser. 2017. Phased genotyping-by-sequencing enhances analysis of genetic diversity and reveals divergent copy number variants in maize. Genes Genomes Genetics In press.
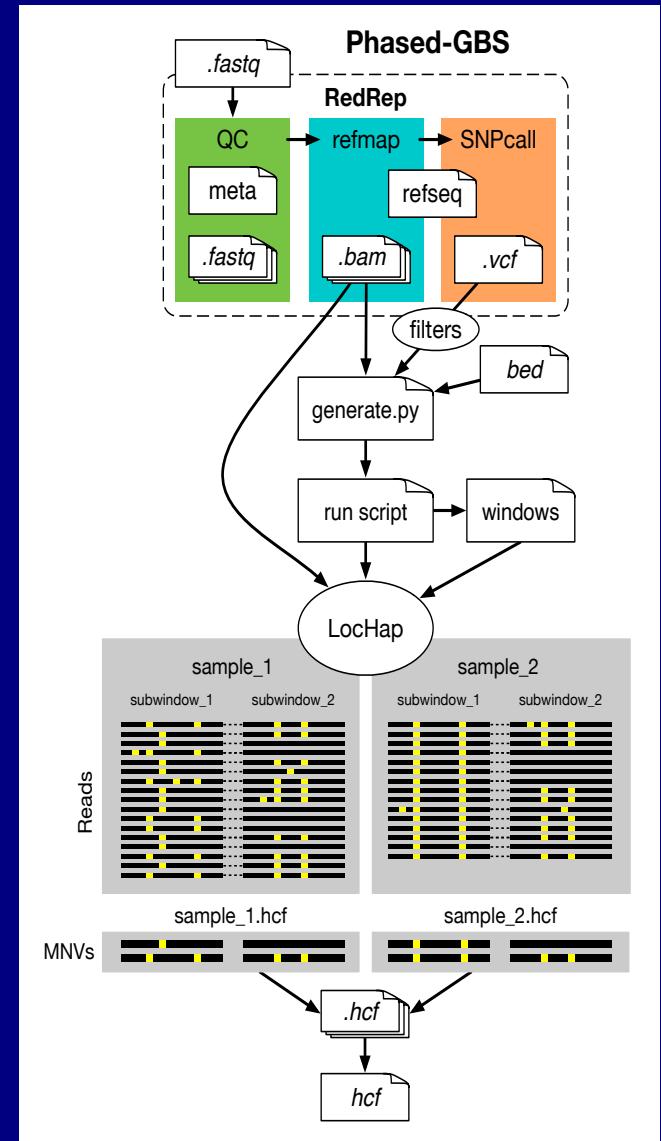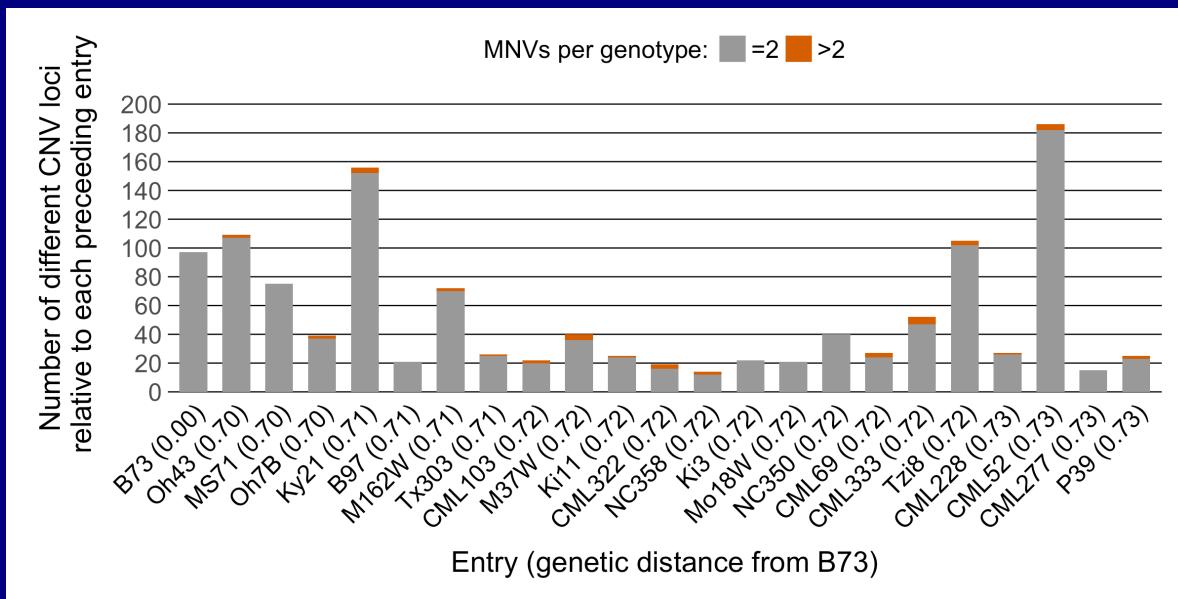
LOD profiles and QTL locations in *A. atriplicis* (R/qtl*)
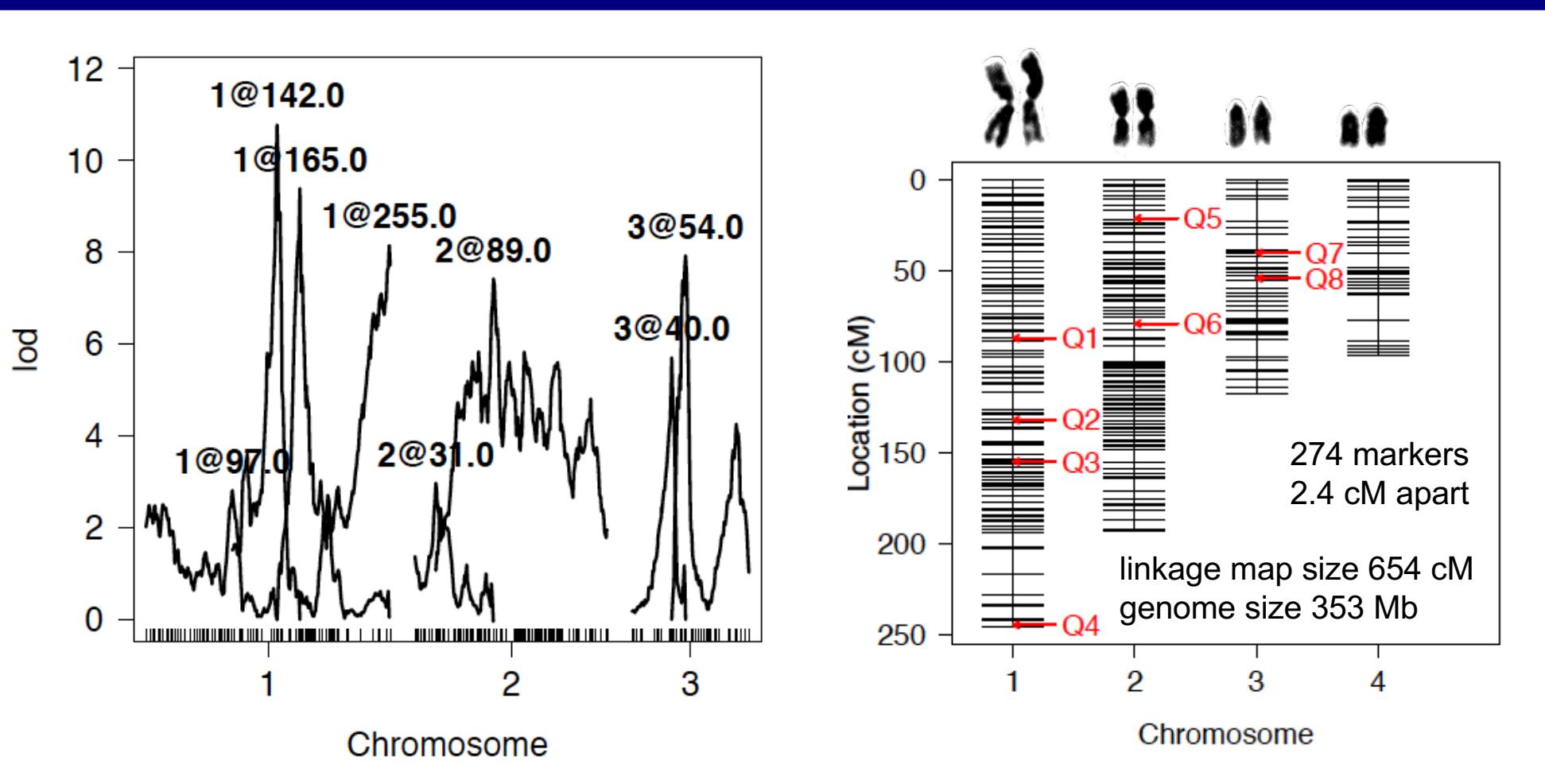
Sizes of effects of QTL and interactions
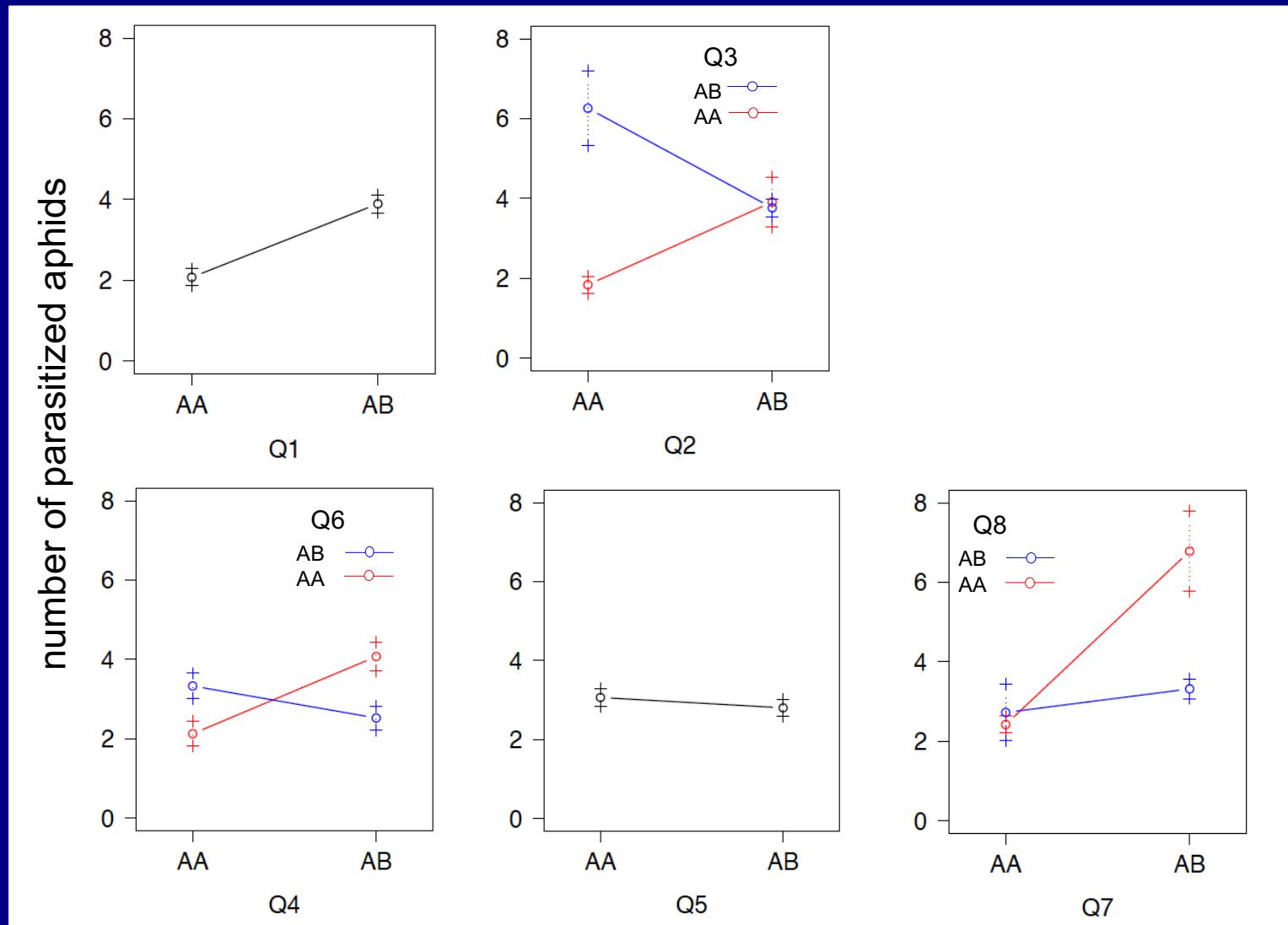
Hopper, K. R., K. L. Kuhn, Q. Li, S. J. Oppenheim, R. J. Wisser, and S. W. Polson. 2015. Genetic architecture of parasitism of a novel host species by a parasitic wasp. Evolution: in review.

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

ANOVA for model:

	df	F	Pr(F)	% variance explained	LOD
Model	11	17.6	<0.00001	39	33.9
Error	308	4.7			

ANOVA for individual QTL and interactions:

	df	F	Pr(F)	% variance explained	LOD	95% CI (cM)
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)
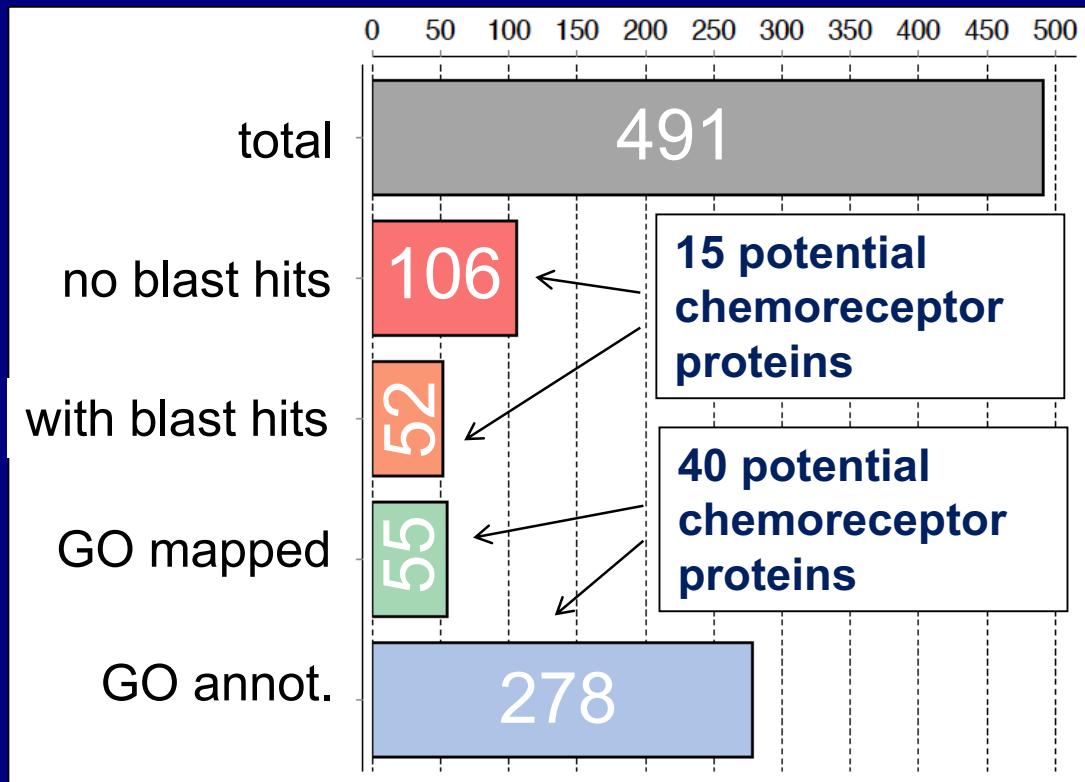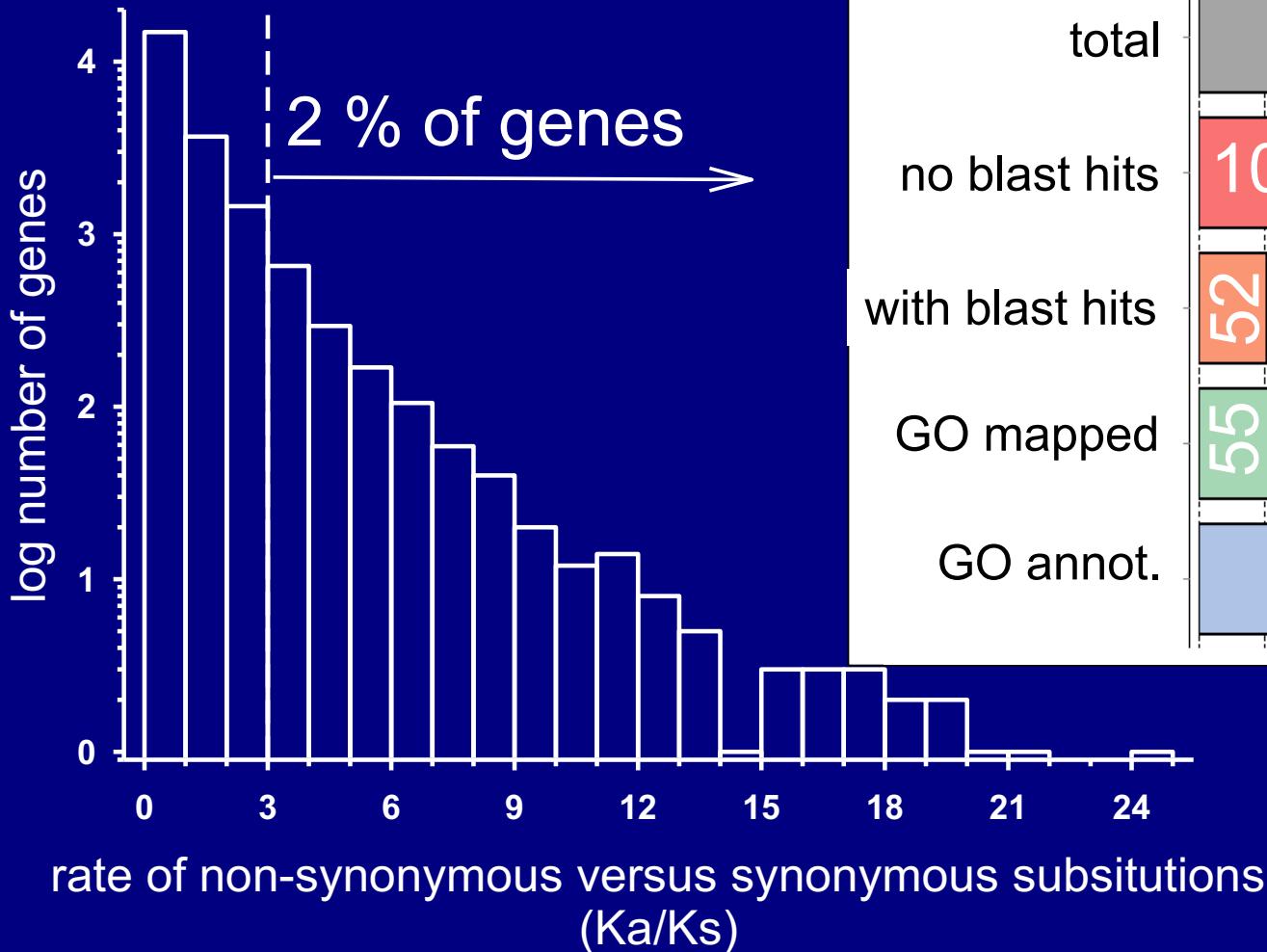
Annotation statistics for genes that are divergent in sequence or expression between *A. atriplicis* and *A. certus*

Sequence divergent genes (Ka/Ks ≥ 3) expressed only in females
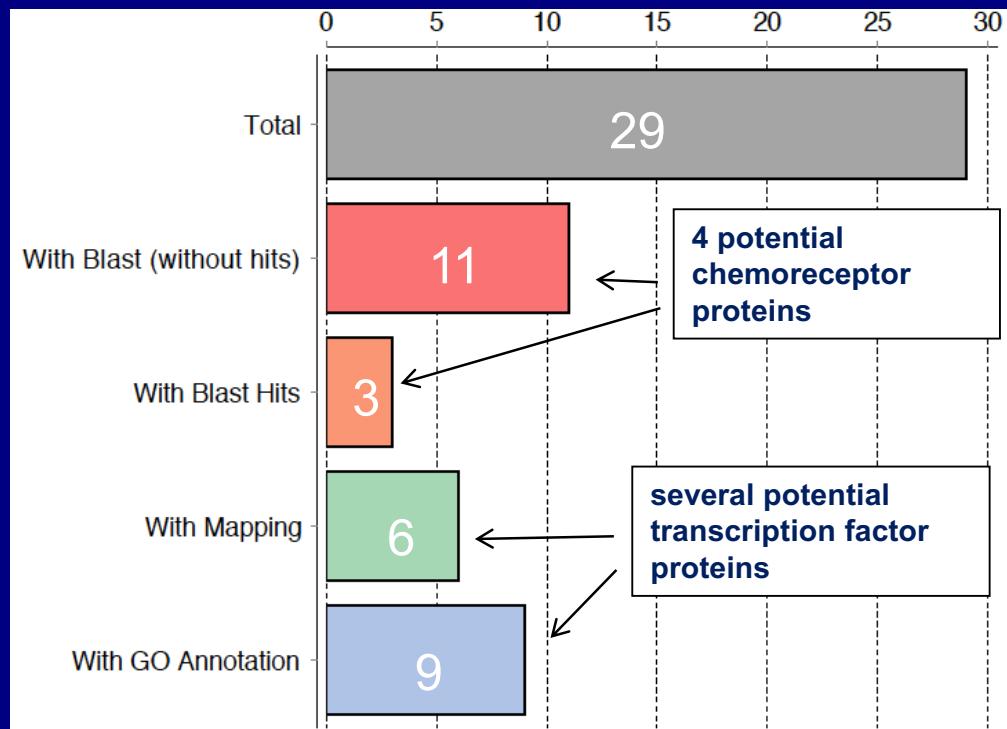
Expression divergent genes (≥ 100x) in females
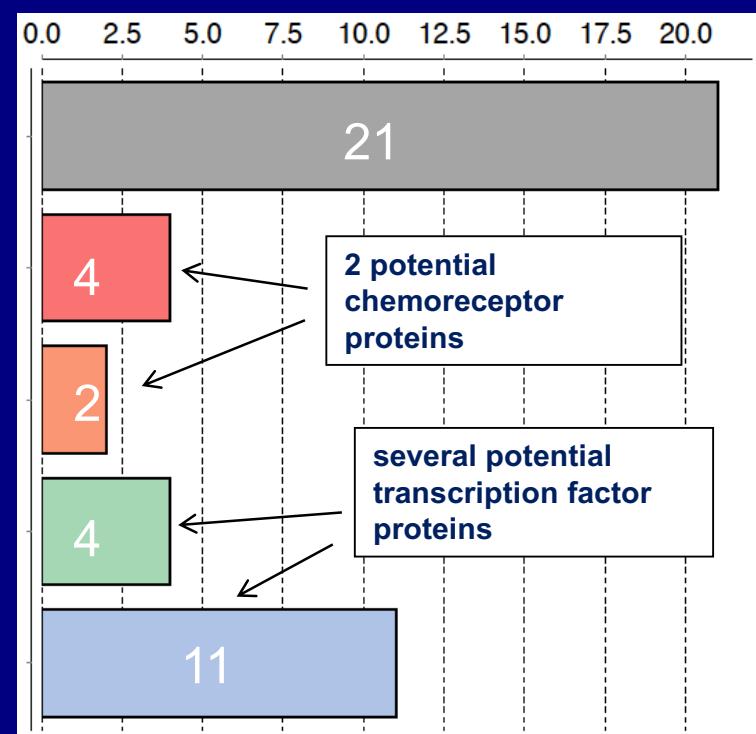
# Chromosomal Fluorescent In Situ Hybridizaton (FISH) to determine gene locations

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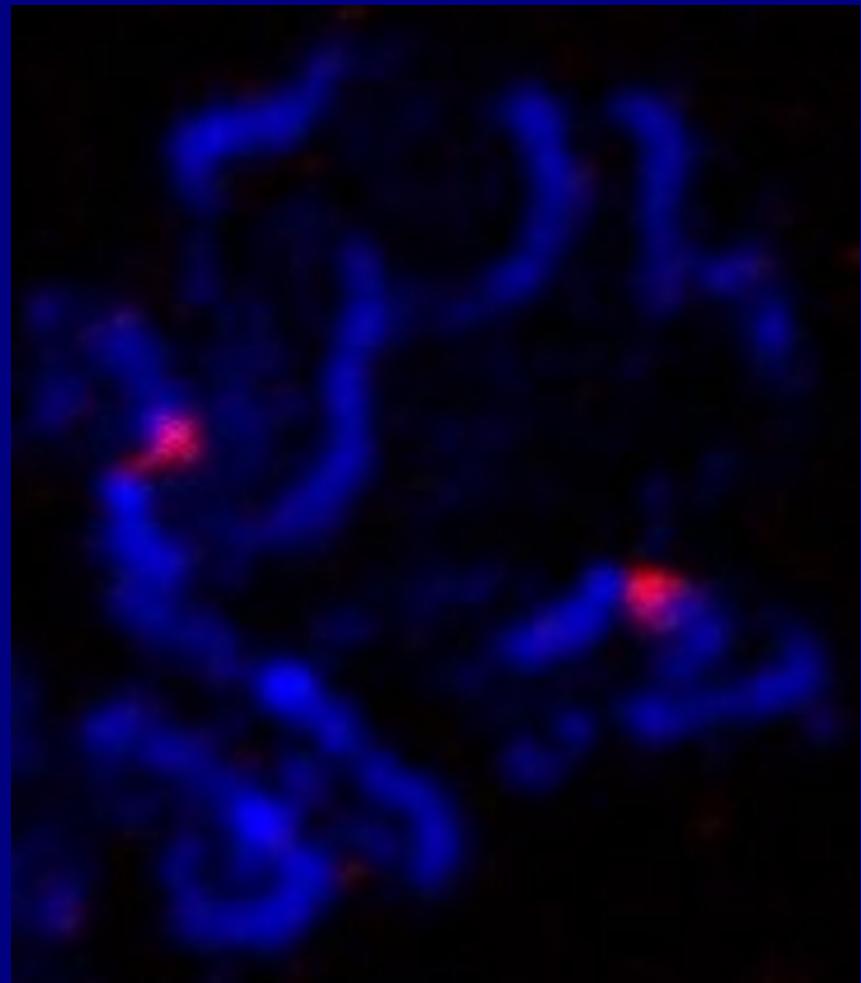
Genotypic candidate genes with Kompetitive Allele Specific PCR

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***

gene 16077 - serine protease precursor
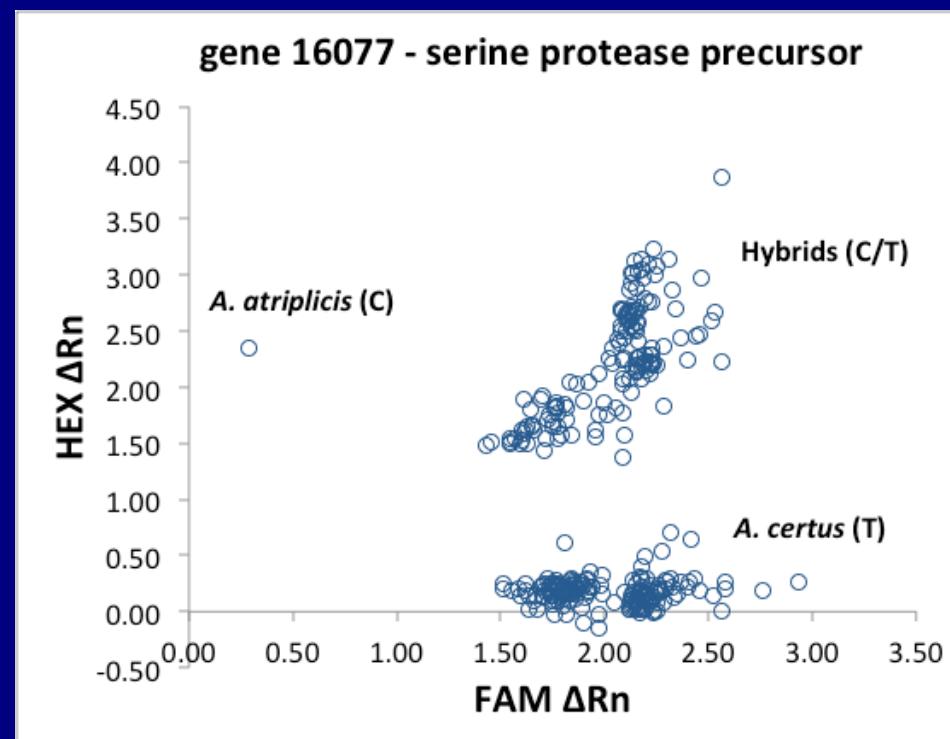
Effect sizes and locations of divergent genes affecting parasitism of *D. noxia* (Blast, InterproScan, R/qtl)

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

* inferred from domain analysis with InterProScan in Blast2GO

Tissue-specific expression of candidate genes

Scanning electron micrograph club of female antenna
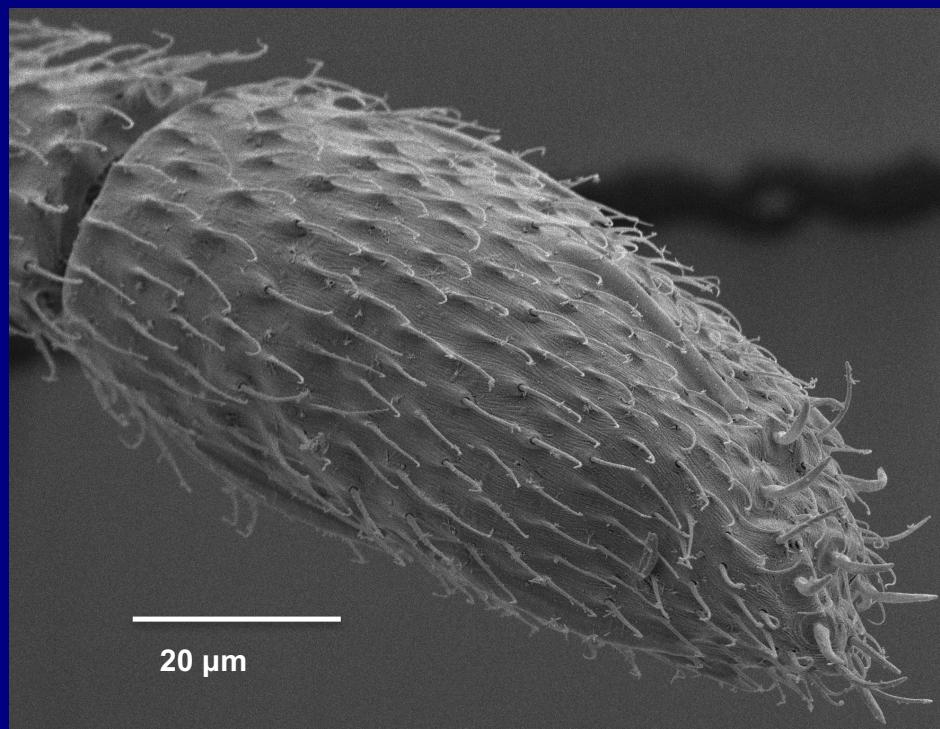
Confocal image of sensilla labelled with red fluorescent probe for mRNA of candidate chemoreceptor gene
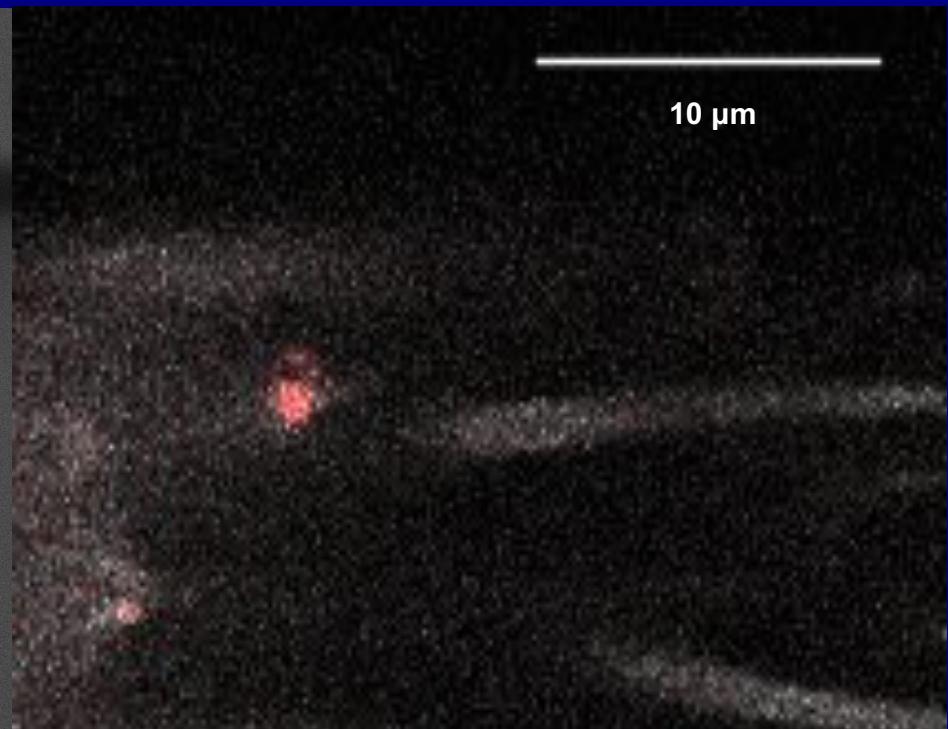
New NIFA-AFRI Grant

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)

Crosses and inbreeding for recombinant inbred lines of *A. atriplicis* x *A. certus* for Multi-parent Advanced Generation InterCross:

♂ *A. atriplicis* x ♀ *A. certus*

X   X   X   X

G_1
G_2
G_3
G_4
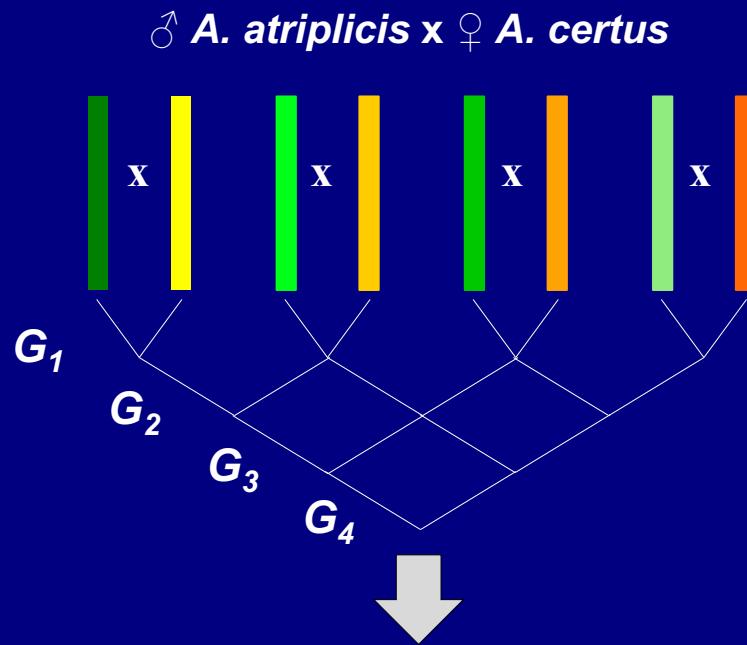
10 generations of brother-sister mating with 400 recombinant inbred lines