

# Two new genome assemblies from aphid parasitoids with low GC content

Slides available: [adennis5.wordpress.com](http://adennis5.wordpress.com)

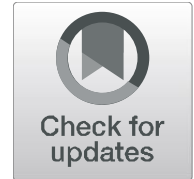
*Photos and screenshots here are ok*


# Work of a large team

RESEARCH ARTICLE

Open Access

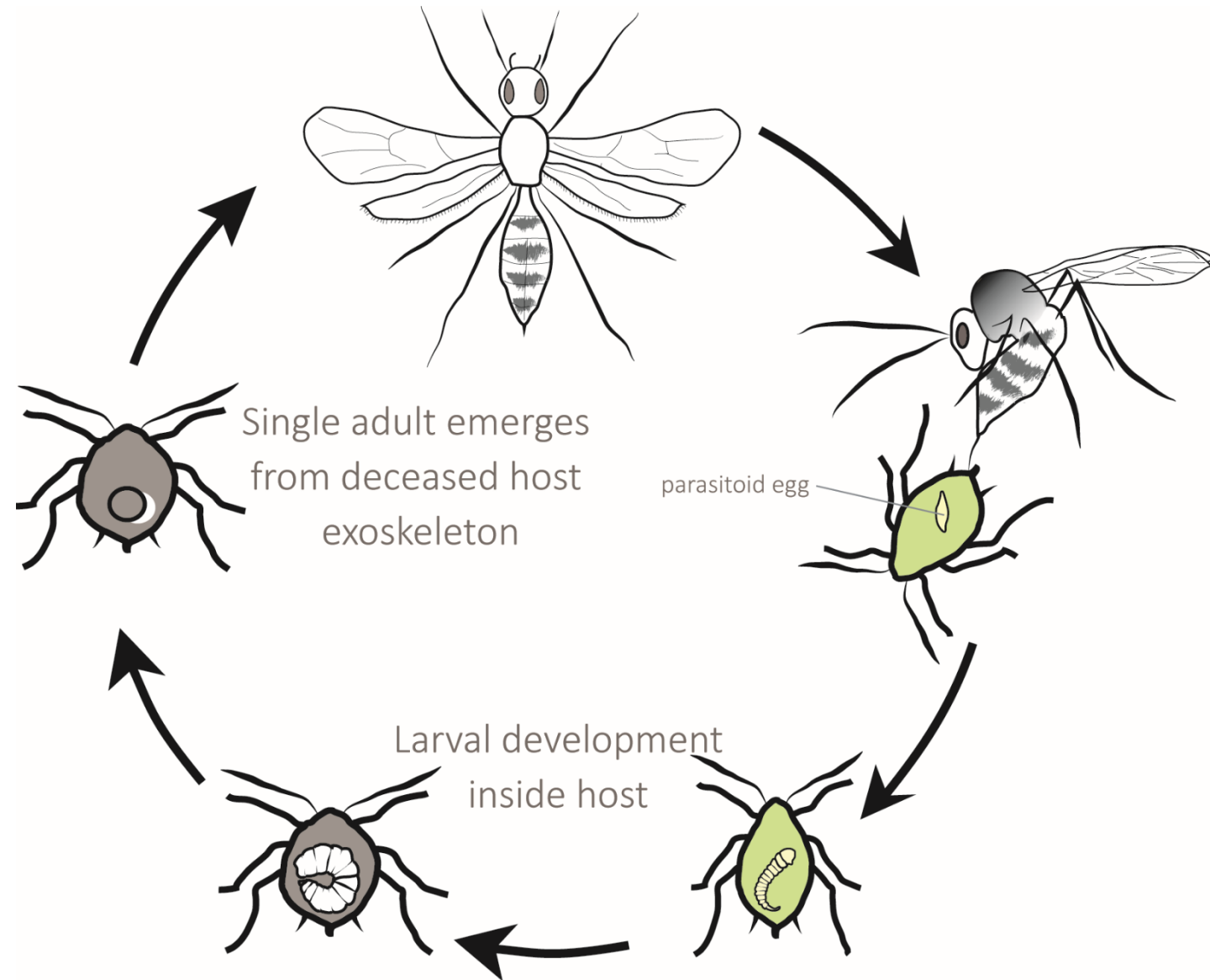
## Functional insights from the GC-poor genomes of two aphid parasitoids, *Aphidius ervi* and *Lysiphlebus fabarum*



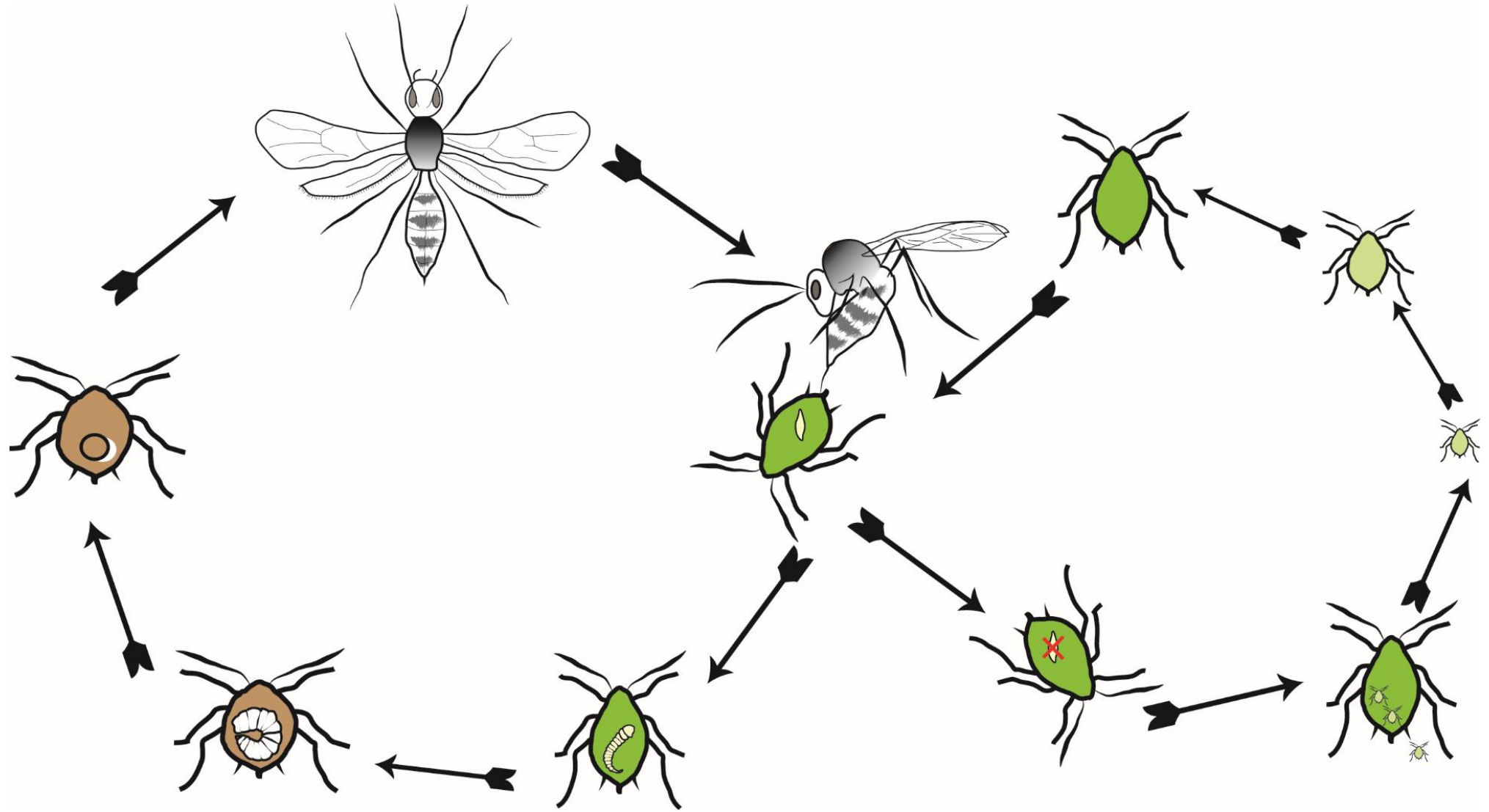
Alice B. Dennis<sup>1,2,3\*</sup> , Gabriel I. Ballesteros<sup>4,5,6†</sup>, Stéphanie Robin<sup>7,8</sup>, Lukas Schrader<sup>9</sup>, Jens Bast<sup>10,11</sup>, Jan Berghöfer<sup>9</sup>, Leo W. Beukeboom<sup>12</sup>, Maya Belghazi<sup>13</sup>, Anthony Bretaudeau<sup>7,8</sup>, Jan Buellesbach<sup>9</sup>, Elizabeth Cash<sup>14</sup>, Dominique Colinet<sup>15</sup>, Zoé Dumas<sup>10</sup>, Mohammed Errbi<sup>9</sup>, Patrizia Falabella<sup>16</sup>, Jean-Luc Gattl<sup>15</sup>, Elzemie k Geuverink<sup>12</sup>, Joshua D. Gibson<sup>14,17</sup>, Corinne Hertaeg<sup>1,18</sup>, Stefanie Hartmann<sup>3</sup>, Emmanuelle Jacquin-Joly<sup>9</sup>, Mark Lammers<sup>9</sup>, Blas I. Lavander<sup>6</sup>, Ina Lindenbaum<sup>9</sup>, Lauriane Massardier-Galata<sup>15</sup>, Camille Meslin<sup>19</sup>, Nicolas Montagné<sup>19</sup>, Nina Pak<sup>14</sup>, Marylène Poirié<sup>15</sup>, Rosanna Salvia<sup>16</sup>, Chris R. Smith<sup>20</sup>, Denis Tagu<sup>7</sup>, Sophie Tares<sup>15</sup>, Heiko Vogel<sup>21</sup>, Tanja Schwander<sup>10</sup>, Jean-Christophe Simon<sup>7</sup>, Christian C. Figueroa<sup>4,5</sup>, Christoph Vorburger<sup>1,2</sup>, Fabrice Legeat<sup>7,8</sup> and Jürgen Gadau<sup>9\*</sup>

[Dennis et al BMC Genomics \(2020\) 21:276](#)

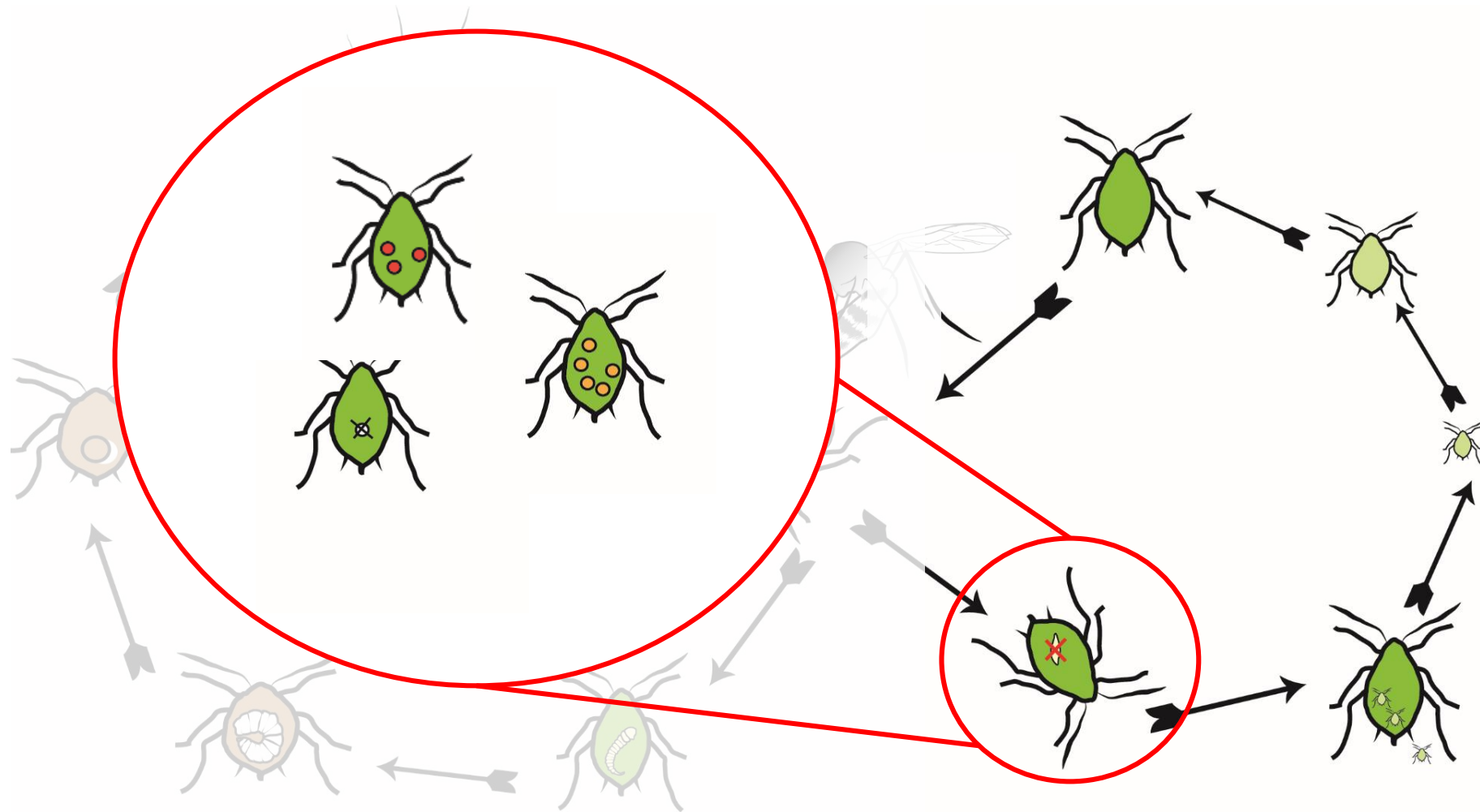
# Aphid parasitoid life cycle



# Strong reciprocal selection



Aphid defense largely from protective endosymbionts  
(but I won't talk about them today)



Why study parasitoids:  
#1 Biocontrol application

# Why study parasitoids: #1 Biocontrol application



LOG

[Fly Eliminators](#) [Beneficial Insects](#) [Pest Control](#) [Disease Control](#) [Grow Supplies](#) [Crop Type](#) [Shop Brands](#)

[Home](#) / [Beneficial Insects](#) / [Complete List Of Predators & Parasites](#) / [Aphidius ervi](#)



## Aphidius ervi

*Aphid Parasite*

Use this aphid parasite to prevent outbreaks of larger aphid species in

**Overnight Shipping Required. Orders ship on Tuesdays Only Tuesday. See Shipping Info for additional information.**

★★★★★ [WRITE A REVIEW](#)

**250 Count**  
SKU: 1001001

\$65.00

1

**BUY NOW**

**500 Count**  
SKU: 1001002

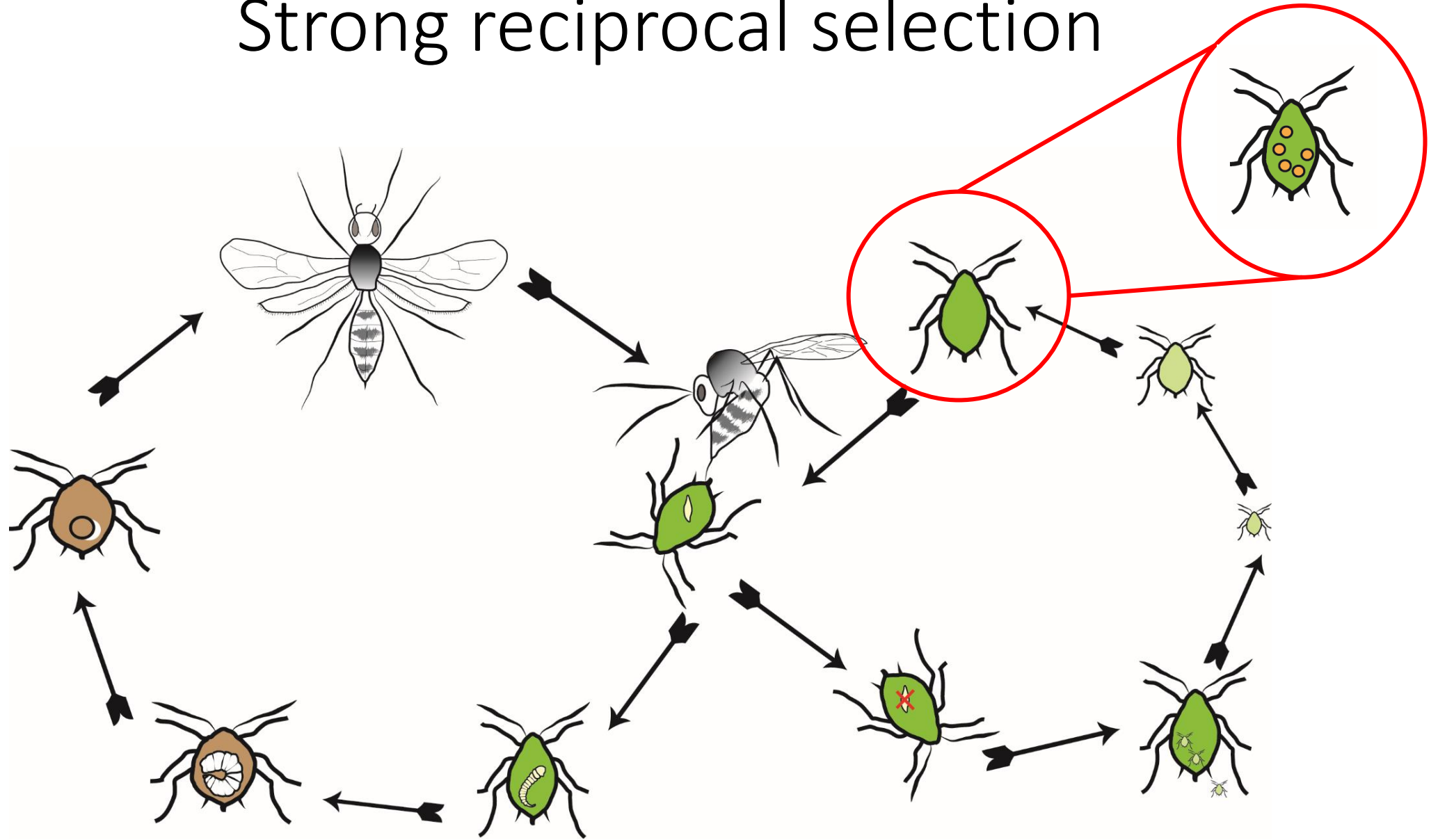
\$100.00

1

**BUY NOW**



# Strong reciprocal selection



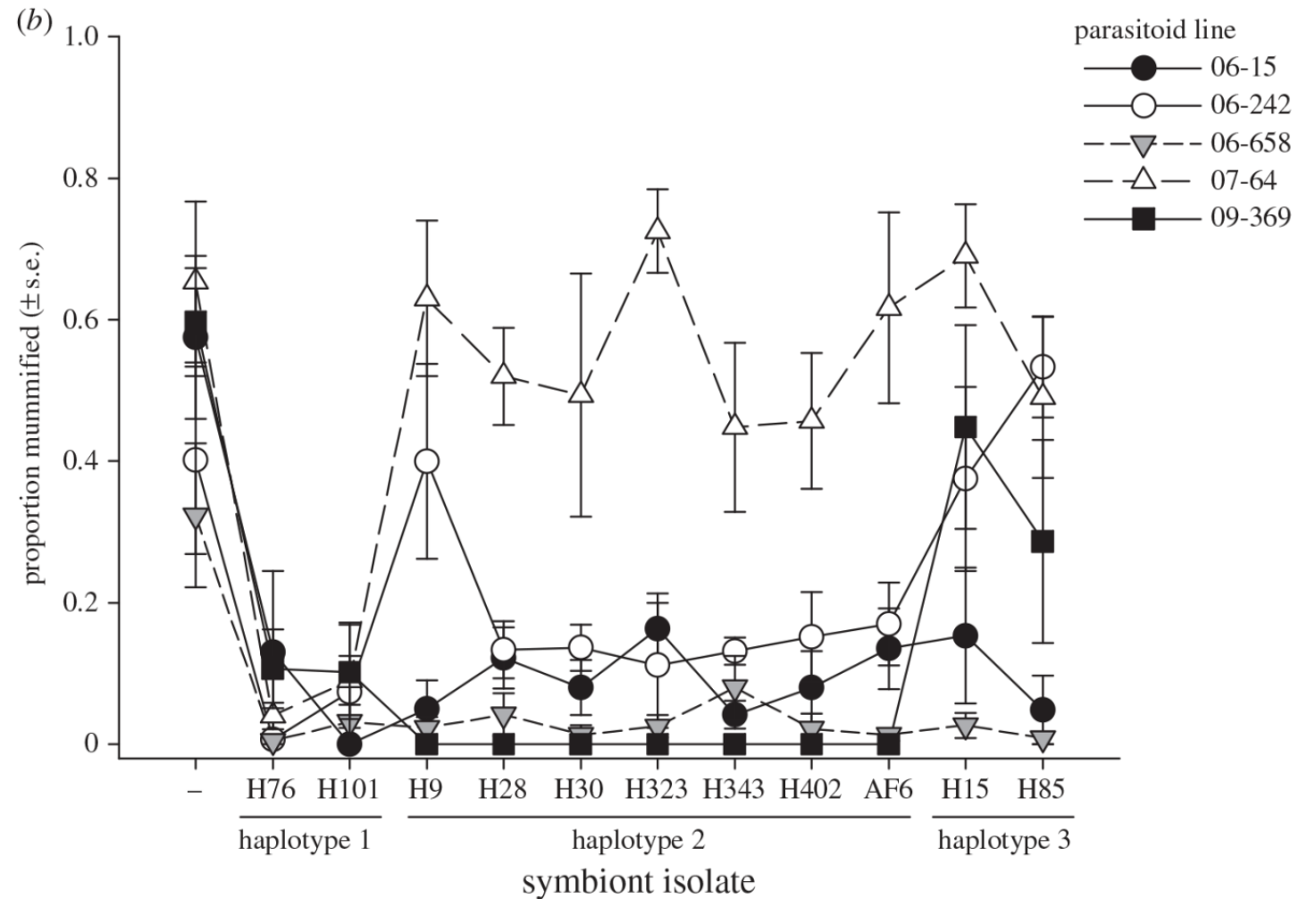


Why study parasitoids:  
#2 Coevolutionary dynamics

# Why study parasitoids: #2 Coevolutionary dynamics

There is variation in parasitoid success across different hosts

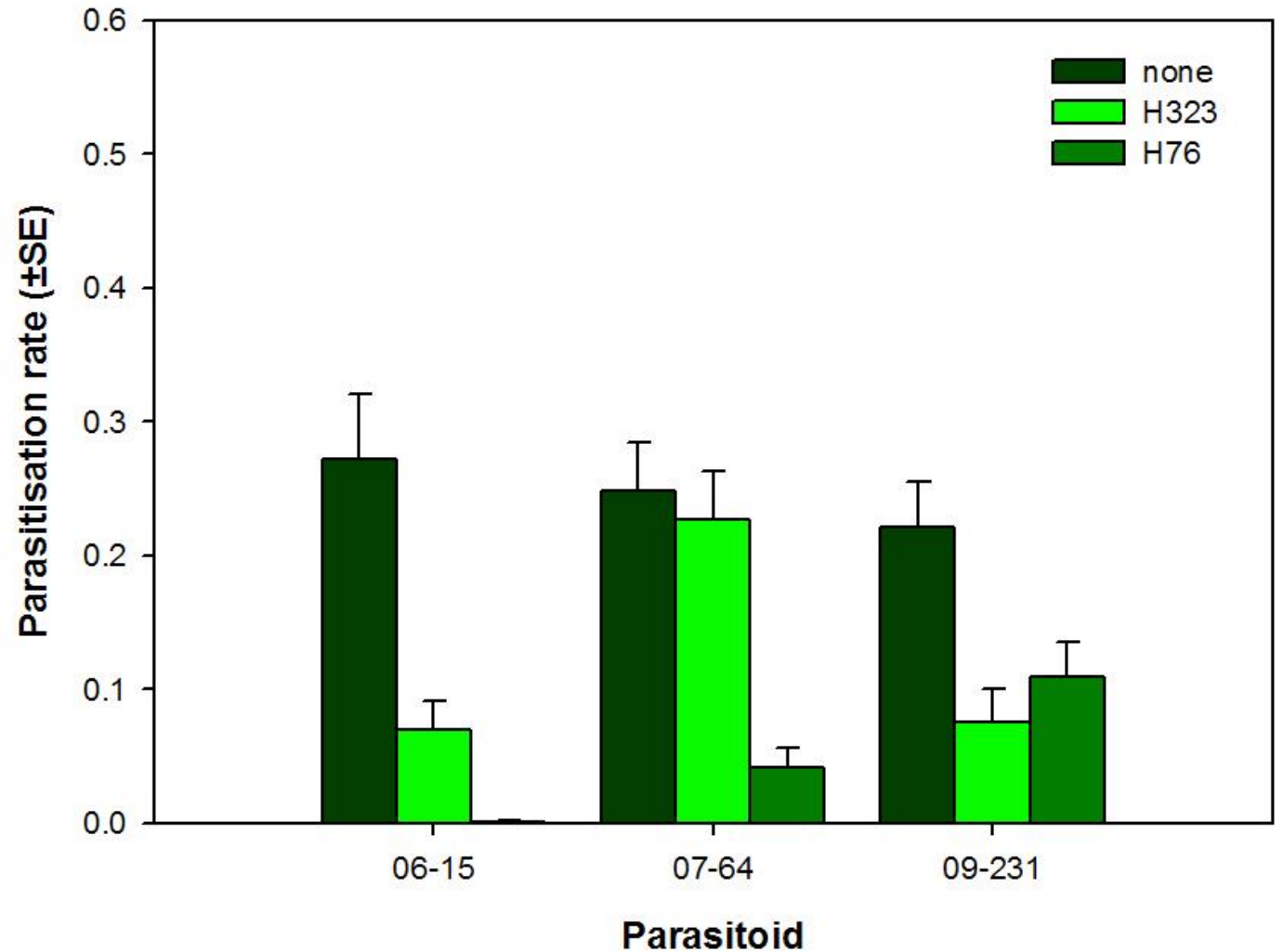
Parasitoid  
*Lysiphlebus fabarum*  
+  
Black bean aphid  
(*A. fabae*) containing  
*H. defensa*



# Why study parasitoids: #2 Coevolutionary dynamics

No universally “best”  
parasitoid

Parasitoid  
*Lysiphlebus fabarum*  
+  
Black bean aphid  
(*A. fabae*) containing  
*H. defensa*



# Genomic resources to facilitate study

Two study species:

*Aphidius ervi*

*Lysiphlebus fabarum*



photograph by Brian Valentine

# Life-history traits

	<b><i>Aphidius ervi</i></b>	<b><i>Lysiphlebus fabarum</i></b>
Host	Aphididae	Aphididae
Primary host tribe	Macrosiphini	Palaeartic
Reproduction	Only sexual	Asexual or sexual
Host is ant tended	No	Yes

# Two different assembly approaches

*Aphidius ervi* data:

Illumina paired end

Illumina Mate Pair (3kb, 5kb, 8kb)

Long read (PacBio)

# Two different assembly approaches

*Aphidius ervi* data:

Illumina paired end

Illumina mate Pair (3kb, 5kb, 8kb)

Long read (PacBio)

*Aphidius ervi* assembly

1. Illumina assembly with:

Platanus + Gap Closer

2. PacBio scaffolding with:

SSPACE + Gap Closer



# Two different assembly approaches

*Aphidius ervi* data:

Paired end

Mate Pair (3kb, 5kb, 8kb)

Long read (PacBio)

*Lysiphlebus fabarum*

Paired end

Mate Pair (5 kb)

Long read (Pac bio)

**Only 1 insert size-** spans smaller parts of genome

**Insert sizes not reliable**

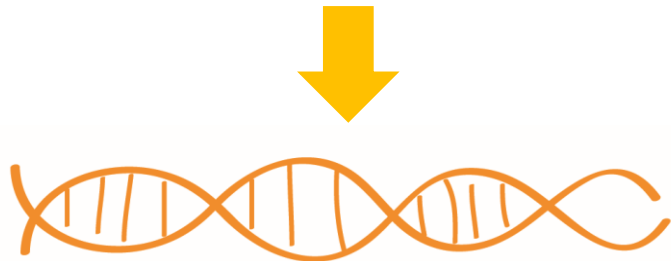
- Determined by mapping raw data to different assemblies
- Checking direction of read mapping
- Issue comes from library preparation



# Two different assembly approaches

## *Lysiphlebus fabarum* assembly

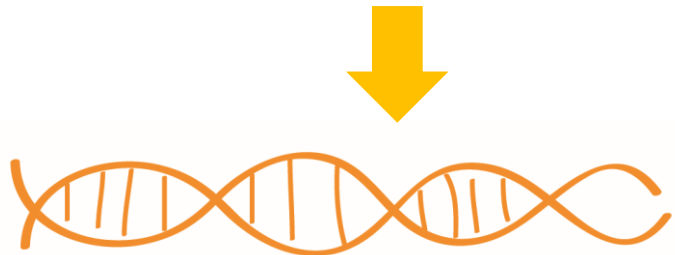
1. MP data filtered for orientation (**Nextclip**)
2. PacBio data error corrected with all Illumina reads (**Proovread**)
3. Long read assembly (**canu**) + polishing (**Pilon**)



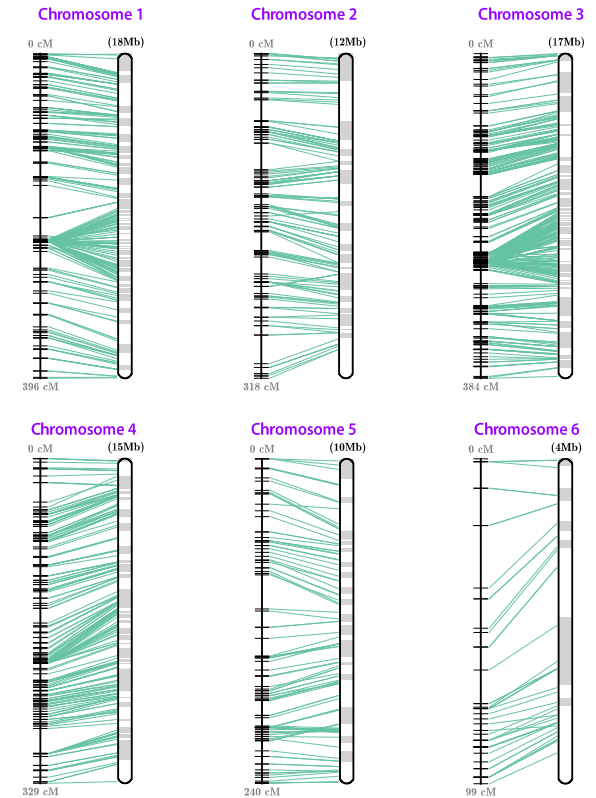
# Two different assembly approaches

## *Lysiphlebus fabarum* assembly

1. MP data filtered for orientation (Nextclip)
2. PacBio data error corrected with all Illumina reads (Proovread)
3. Long read assembly (canu) + polishing (Pilon)



Predicted chromosomes for *Lysiphlebus fabarum*



+ linkage mapping from  
haploid sons of a single  
mom

# Results

	<i>Aphidius ervi</i>	<i>Lysiphlebus fabarum</i>
Total length	139 Mbp	141 Mbp
Longest scaffold	3.7 Mbp	2.2 Mbp
N scaffolds / contigs	5,743	1,698
N50	581,355 bp	216,143 bp
GC %	25.8	23.8

# Consequences of assembly differences?

	<i>Aphidius ervi</i>	<i>Lysiphlebus fabarum</i>
Total length	139 Mbp	141 Mbp
Longest scaffold	3,671,467 bp	2,183,677 bp
N scaffolds / contigs	5,743	1,698
N50	581,355 bp	216,143 bp
GC %	25.8	23.8
N scaffolds / contigs > 3,000 bp	<b>1,503</b>	<b>1,698</b>
Quality differences	BUSCO overall better	more duplication

# Predictive annotation with MAKER



	<i>Aphidius ervi</i>	<i>Lysiphlebus fabarum</i>
Predicted genes	20,344	15,203
Genome coverage by CDS	17.8%	14.9%
CDS with transcriptome support	77.8%	88.3%

# Annotation by community

	<i>Aphidius ervi</i>	<i>Lysiphlebus fabarum</i>
Predicted genes	20,344	15,203

# Annotation by community

Online Apollo. Accessed by many researchers.

1 scaffold

**User edits**

**Maker annotations**

**RNA-seq reads**

Annotations

AE3001583-RA-00001

AE3001579-RA

AE3001581-RA

AE3001580-RA

AE3001582-RA

AE3001583-RA

Tissue : Body, Host : A.pisum, Reared on : Pea (BFAPP)

French strain : Venom apparatus - SNPs/Coverag

Tissue : Body, Host : A.pisum, Reared on : alfalfa

Tissue : Body, Host : A.pisum, Reared on : alfalfa

Tissue : Body, Host : A.pisum, Reared on : Pea (B

Tissue : Body, Host : A.pisum, Reared on : Pea (B

Tissue : Body, Host : S. avenae (BS2SA)

Details

Name

Type

Configuration

# Annotation: the results of many hours

Category	<i>A. ervi</i>	<i>L. fabarum</i>
Venom proteins	32	35
Desaturases	14	11
Immune genes	270	264
Osiris genes	21	25
OXPPOS	75	74
Odorant receptors (ORs)	228	156
Ionotropic receptors (IRs)	42	40
Odorant-binding proteins (OBPs)	14	11
Chemosensory proteins (CSPs)	11	13
Core (transformer, doublesex)	4	3
Sex related genes	6	5
DNA Methylation genes	2	2
<b>TOTALS</b>	<b>719</b>	<b>642</b>

Chemosensory

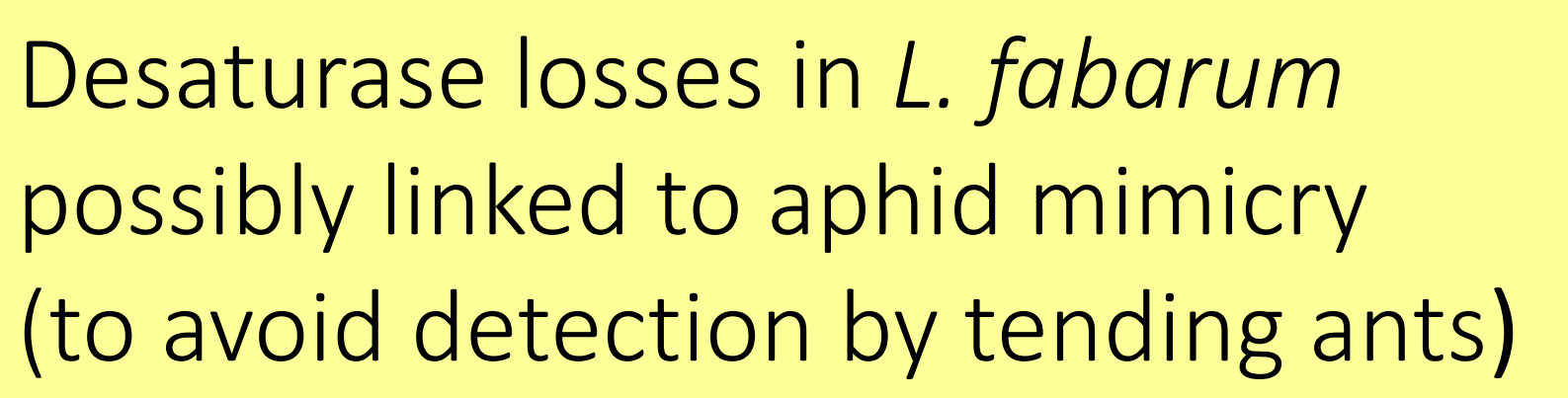
Sex



# Annotation: the results of many hours

Category	<i>A. ervi</i>	<i>L. fabarum</i>
Venom proteins	32	35
Desaturases	14	11
Immune genes	270	264
Other	1	25
Other	5	74
Other	8	156
Other	2	40
Other	4	11
Chemosensory proteins (CSPs)	11	13
Core (transformer, doublesex)	4	3
Sex related genes	6	5
DNA Methylation genes	2	2
TOTALS	719	642

Desaturase losses in *L. fabarum* possibly linked to aphid mimicry (to avoid detection by tending ants)



Sex

# Annotation: the results of many hours

Category	<i>A. ervi</i>	<i>L. fabarum</i>
Venom proteins	32	35
Desaturases	14	11
Immune genes	270	264
Osiris genes	21	25
OXPPOS	75	74
Odorant receptors (ORs)		156
Ionotropic receptors (IRs)		40
Odorant-binding proteins (OBPs)		11
Chemosensory proteins (CSPs)	11	13
Core (transformer, doublesex)	4	3
Sex related genes	6	5
DNA Methylation genes	2	2
TOTALS	719	642

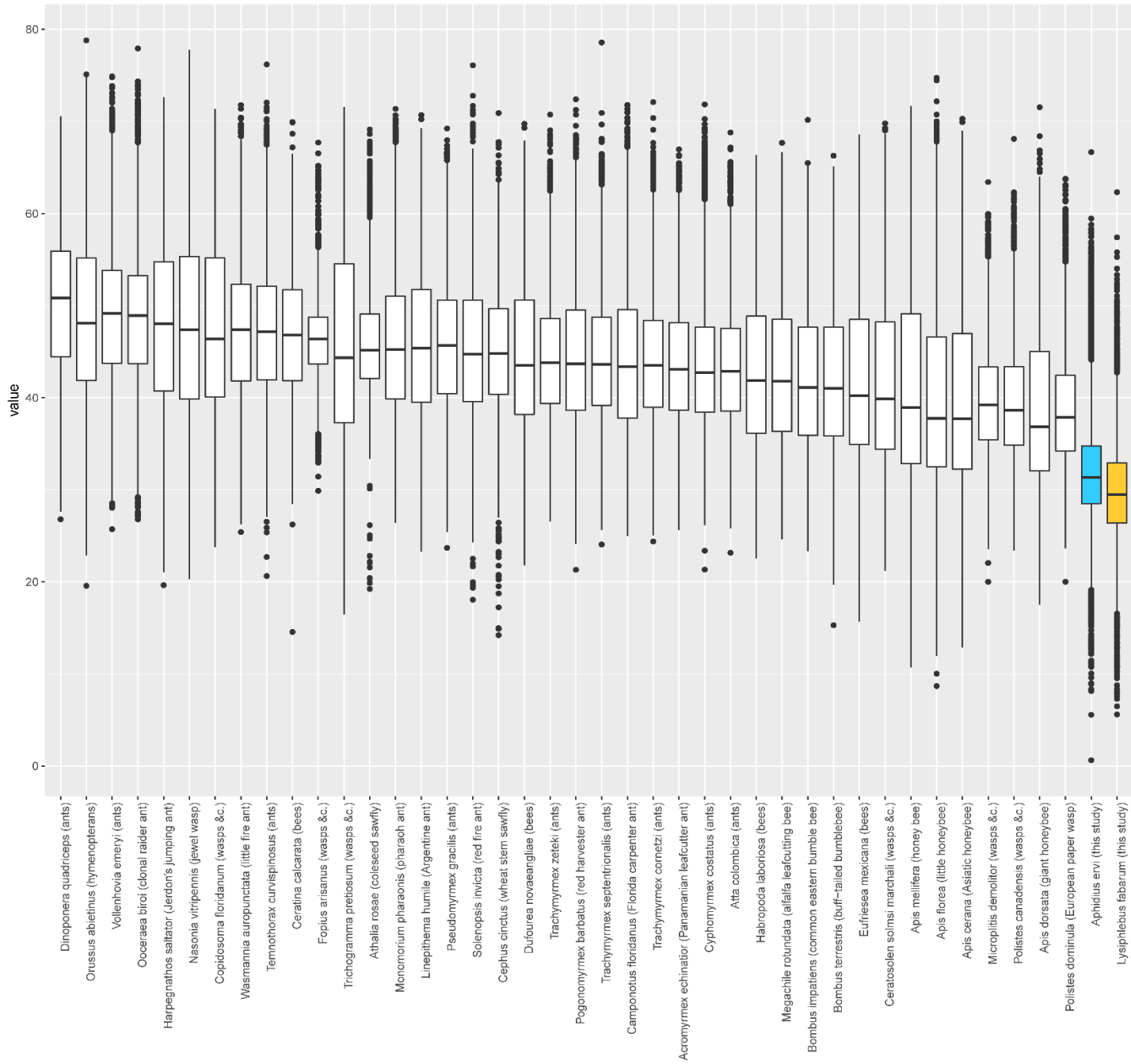
DNA methyltransferase 1  
not found

GC content

# Predictive annotation with MAKER



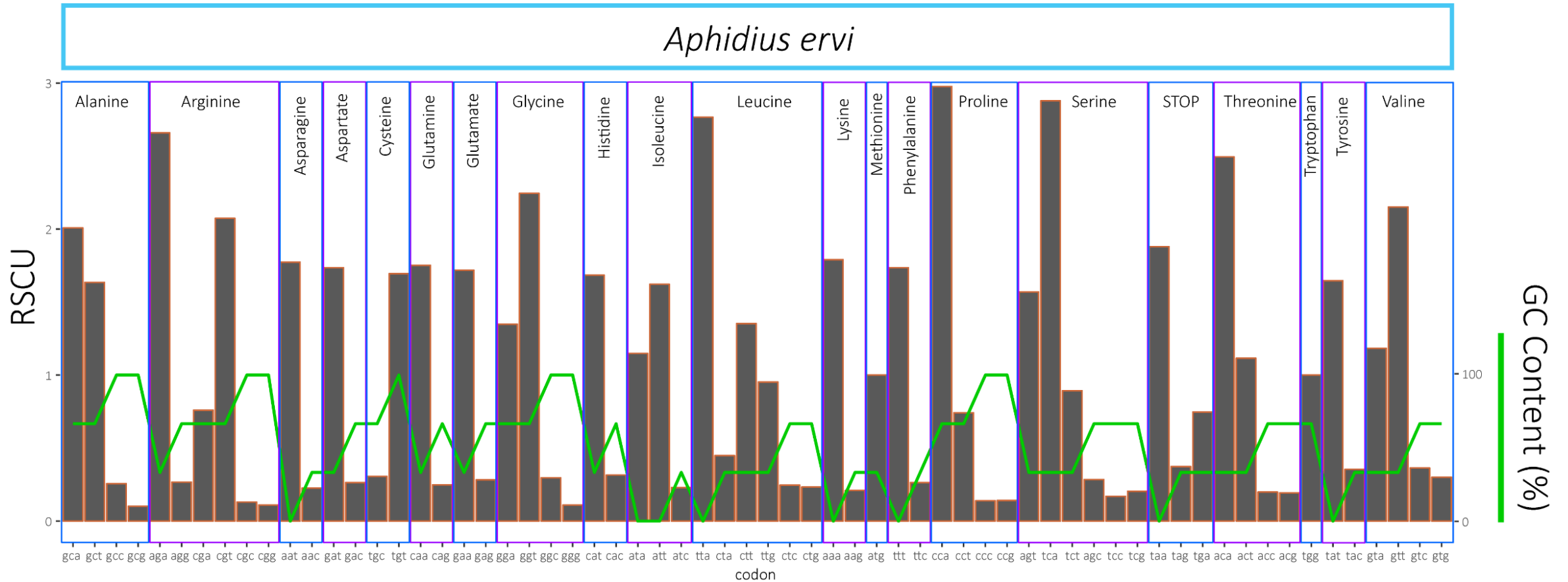
	<i>Aphidius ervi</i>	<i>Lysiphlebus fabarum</i>
Predicted genes	20,344	15,203
Genome coverage by CDS	17.8%	14.9%
CDS with transcriptome support	77.8%	88.3%
GC content of genome	25.8%	23.8%
Mean GC of CDS	31.9%	29.8%
GC of 3 <sup>rd</sup> positions	15.5%	10.7%



GC content

Compared to CDS of  
42 Hymenoptera

# Codon usage favors low GC-codons



A cost to synonymous changes?

# Why such low GC content?

- Small genome size
- Lack of methylation
- Low  $N_e$
- Mutational bias

However, none of these are unique to these taxa.

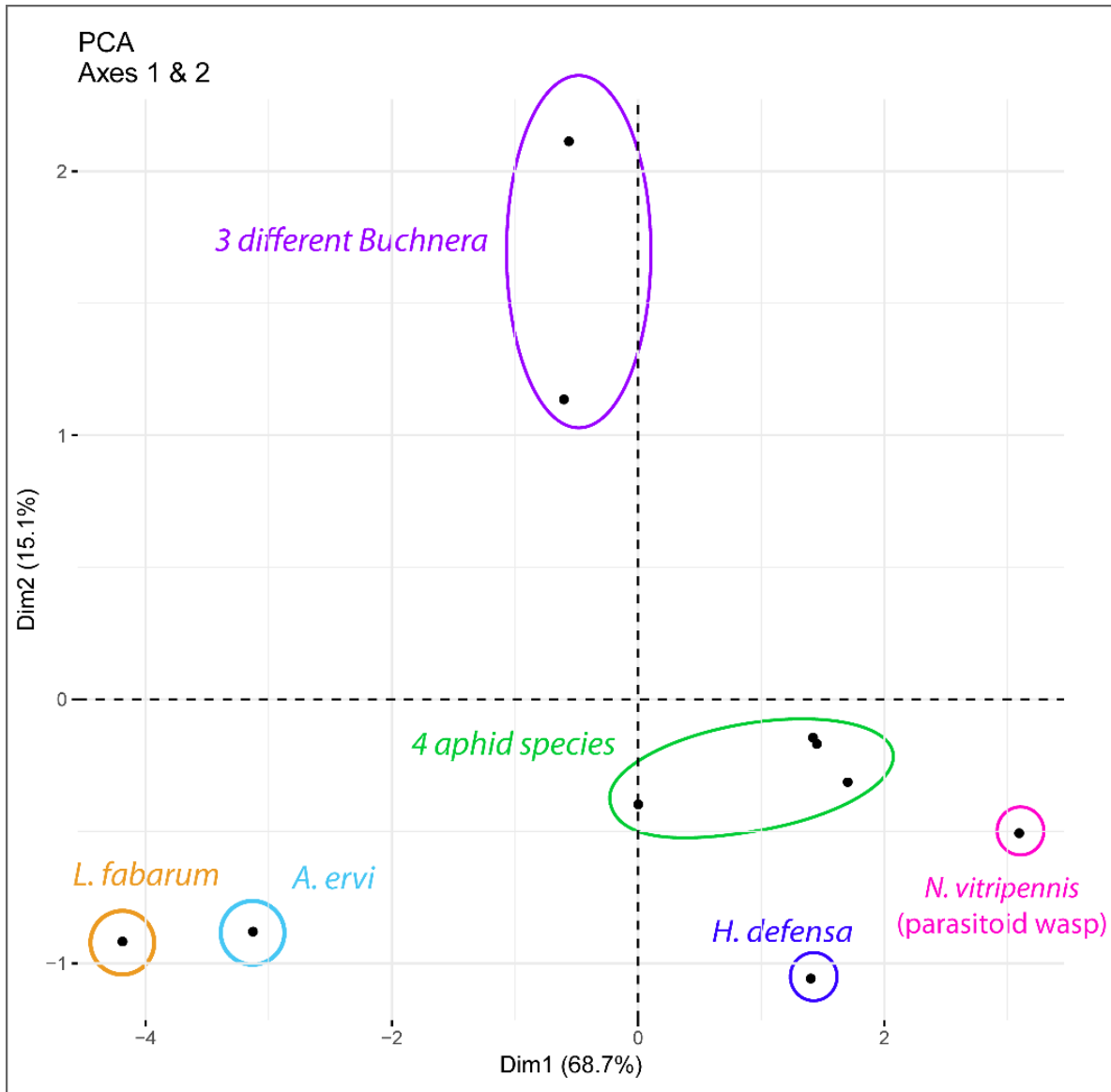
# GC content – idea #1

Host aphids also have low GC content.

Is there similarity in codon usage across taxa in this system?



# GC content – idea #1



Host aphids also have low GC content.

Is there similarity in codon usage across taxa in this system?

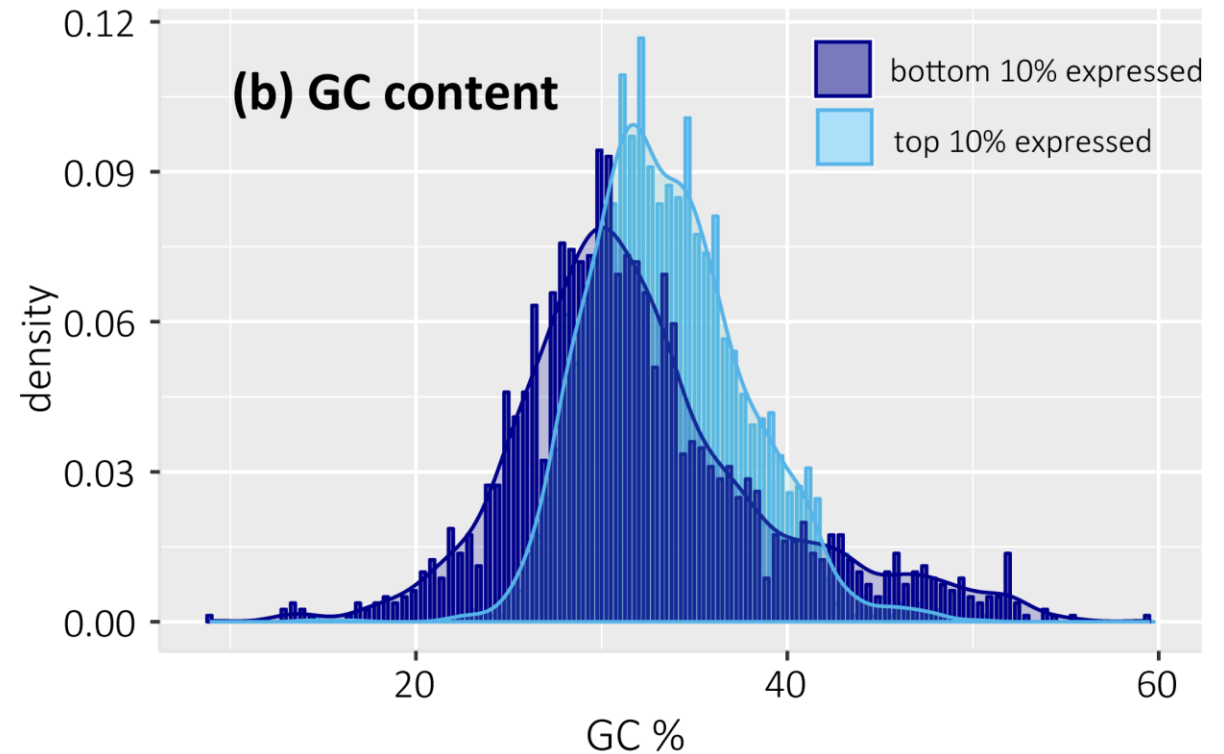
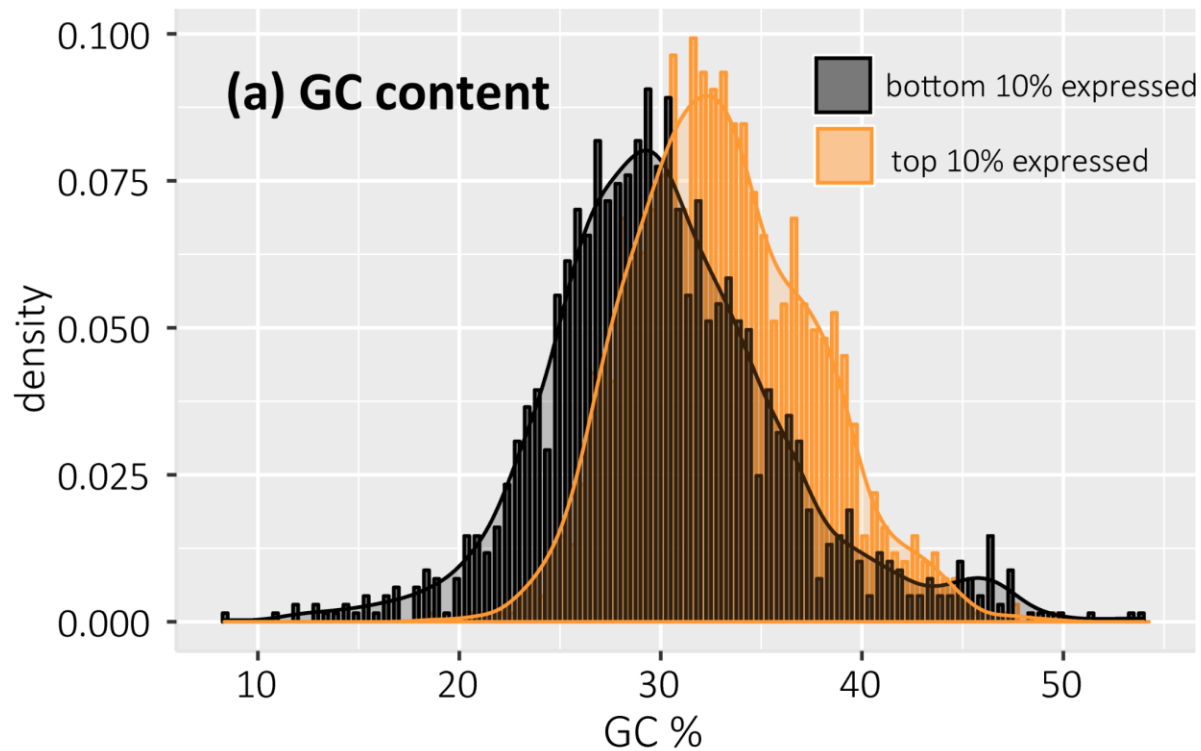
**No.**

## GC content – idea #2

Selection should be strongest on the most highly expressed genes. Do these have lower GC content?

# GC content – idea #2

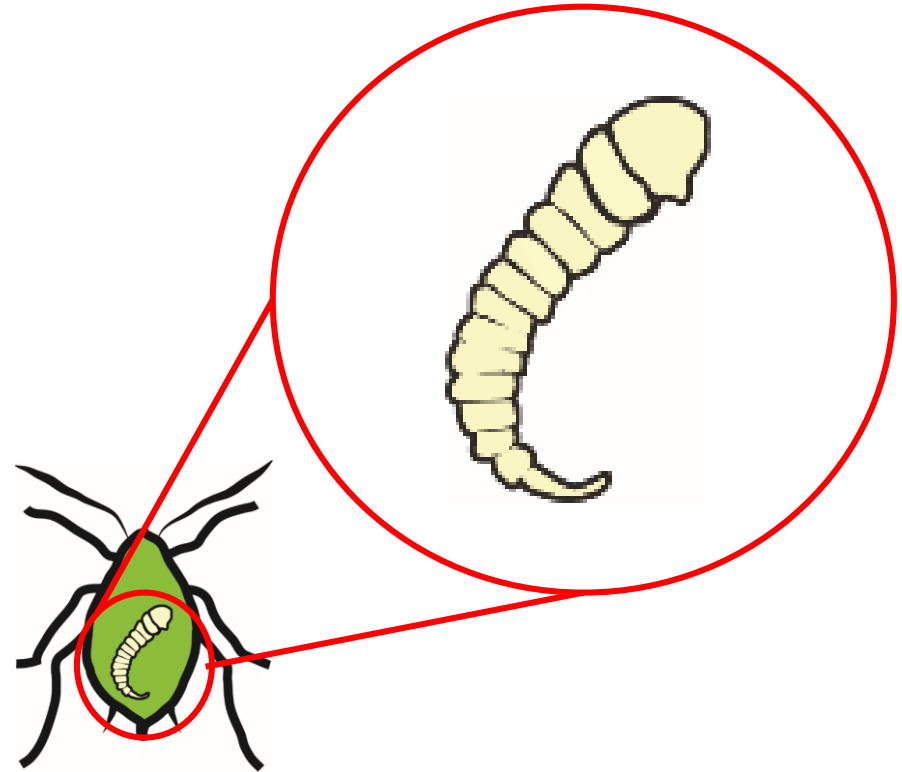
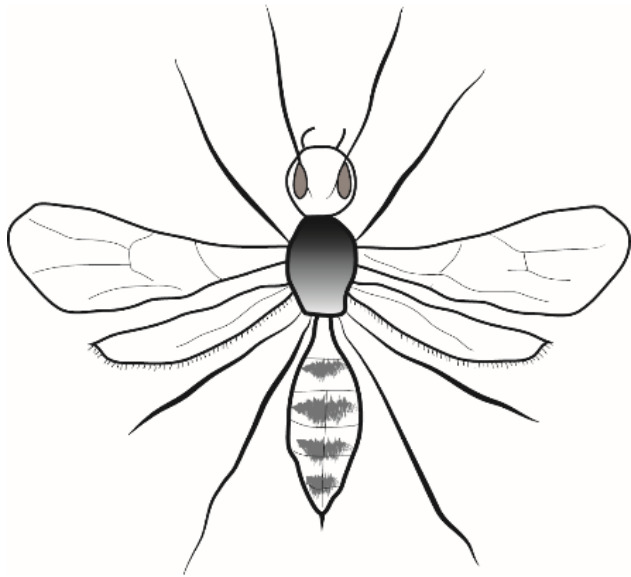
Selection should be strongest on the most highly expressed genes. Do these have lower GC content?



**No. Higher GC content in top expressed genes is in line with observations in other taxa.**

# GC content – idea #3

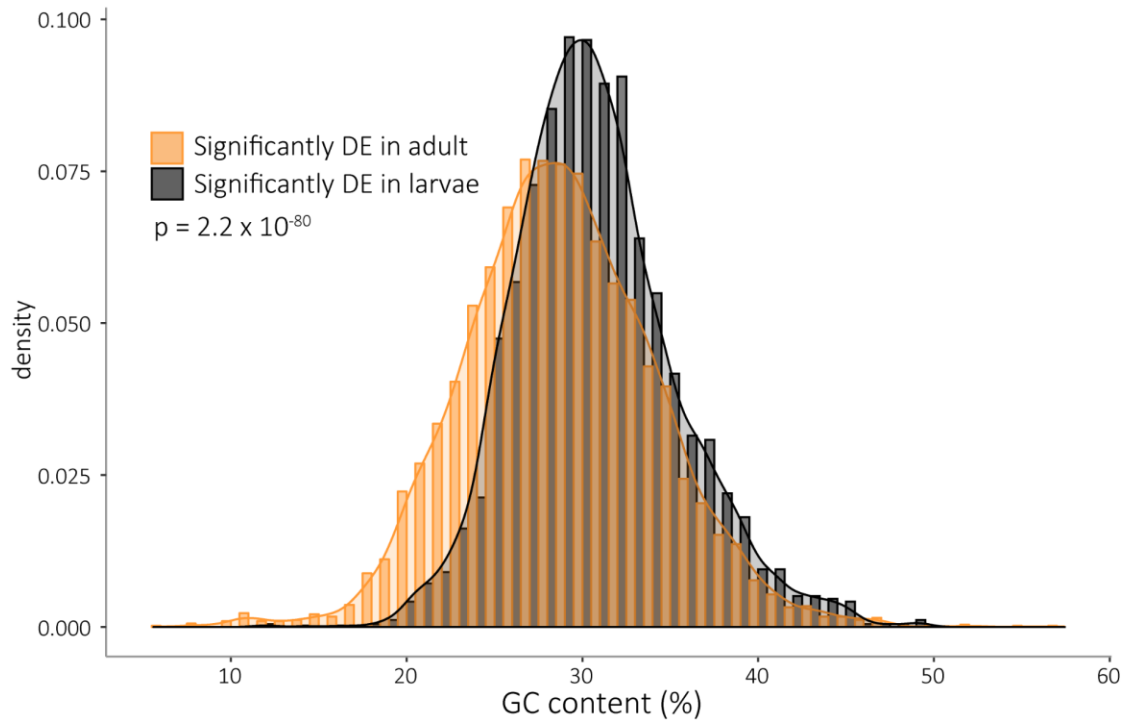
Life- history specific expression



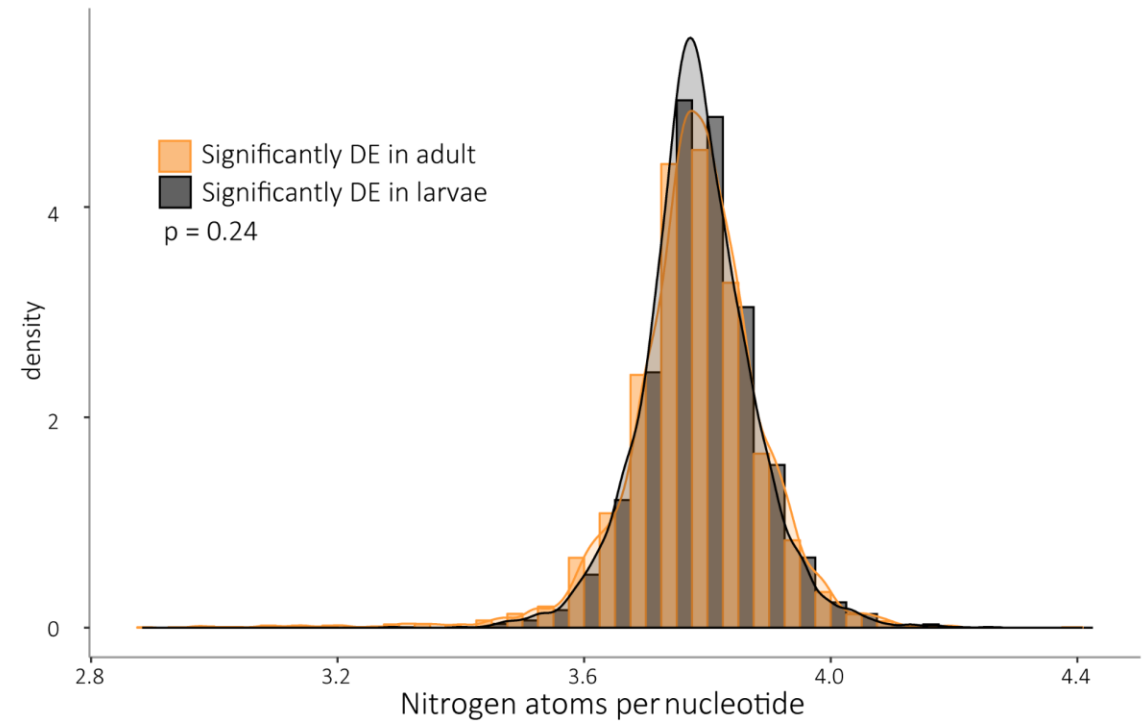
For *L. fabarum* we have RNA-seq from both larvae and adults

# Adult-biased genes have lower GC content

## GC content

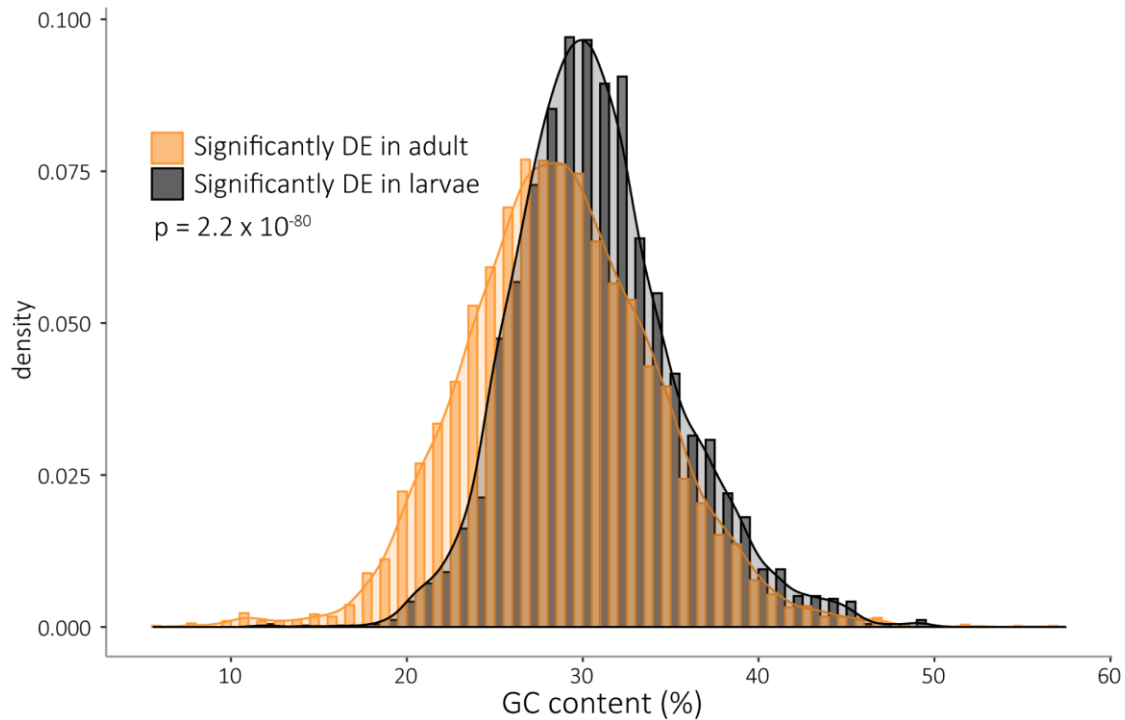


## Nitrogen content



# Adult-biased genes have lower GC content

GC content



Possible explanations:

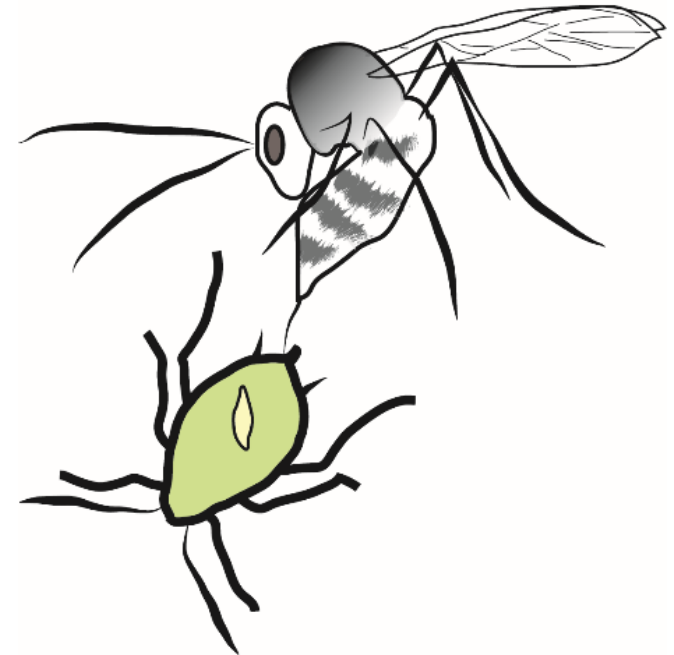
More energy constraints in adults?

Biased gene conversion?

GC3= 15.5% and 10.7%

# Conclusions

- Aphid parasitoids offer chance to study coevolutionary dynamics
- Curation of key gene families key to future work
- Low GC content has evolutionary implications





## Available organisms

*Aphidius ervi*



Other organisms are available on the [BIPAA website](#).

## Available datasets

*Aphidius ervi*

**Structural annotation: Whole Genome Assembly (v3.0) and Annotation (v3.0) of Aphidius ervi**

20226 genes, performed on 2017-01-26

**Functional annotation: Blast2GO: Annotation v3.0[Report]**

5870 annotated proteins, performed on 2017-01-27

**Functional annotation: Blastp: Annotation v3.0 vs NR**

20344 annotated proteins, performed on 2017-01-27

**Functional annotation: InterProScan: Annotation v3.0**

14901 annotated proteins, performed on 2017-01-27

[BIPAA Portal: https://bipaa.genouest.org/is/](https://bipaa.genouest.org/is/)



Thank you!



FONDS NATIONAL SUISSE  
SCHWEIZERISCHER NATIONALFONDS  
FONDO NAZIONALE SVIZZERO  
SWISS NATIONAL SCIENCE FOUNDATION



Deutsche  
Forschungsgemeinschaft



Fondecyt  
Fondo Nacional de Desarrollo  
Científico y Tecnológico



Available to chat  
after today's session  
or by arrangement

Dennis et al. BMC Genomics (2020) 21:376  
<https://doi.org/10.1186/s12864-020-6764-0>

BMC Genomics

RESEARCH ARTICLE

Open Access



## Functional insights from the GC-poor genomes of two aphid parasitoids, *Aphidius ervi* and *Lysiphlebus fabarum*

Alice B. Dennis<sup>1,2,3\*†</sup>, Gabriel I. Ballesteros<sup>4,5,6†</sup>, Stéphanie Robit<sup>7,8</sup>, Lukas Schrader<sup>9</sup>, Jens Bast<sup>10,11</sup>, Jan Berghöfer<sup>9</sup>, Leo W. Beukeboom<sup>12</sup>, Maya Belghazi<sup>3</sup>, Anthony Bretaudeau<sup>8</sup>, Jan Buellesbach<sup>9</sup>, Elizabeth Cash<sup>14</sup>, Dominique Colinet<sup>15</sup>, Zoé Dumas<sup>10</sup>, Mohammed Errbi<sup>8</sup>, Patrizia Falabella<sup>6</sup>, Jean-Luc Gatti<sup>15</sup>, Elzemie Geuverink<sup>2</sup>, Joshua D. Gibson<sup>14,17</sup>, Corinne Hertaeg<sup>1,18</sup>, Stefanie Hartmann<sup>3</sup>, Emmanuelle Jacquin-Joly<sup>9</sup>, Mark Lammers<sup>9</sup>, Blas I. Lavander<sup>6</sup>, Ina Lindenbaum<sup>9</sup>, Lauriane Massardier-Galata<sup>15</sup>, Camille Meslin<sup>9</sup>, Nicolas Montagné<sup>9</sup>, Nina Pak<sup>14</sup>, Marylène Poirié<sup>15</sup>, Rosanna Salviá<sup>6</sup>, Chris R. Smith<sup>20</sup>, Denis Tagu<sup>7</sup>, Sophie Tares<sup>15</sup>, Heiko Vogel<sup>1</sup>, Tanja Schwander<sup>10</sup>, Jean-Christophe Simon<sup>7</sup>, Christian C. Figueroa<sup>4,5</sup>, Christoph Vorburger<sup>1,2</sup>, Fabrice Legeat<sup>7,8</sup> and Jürgen Gadau<sup>9\*</sup>

[Dennis et al BMC Genomics \(2020\) 21:276](#)

[BIPAA Portal: https://bipaa.genouest.org/is/](https://bipaa.genouest.org/is/)

[Slides at: adennis5.wordpress.com](https://adennis5.wordpress.com)

EXTRA SLIDES  
AND INFO

# *A. ervi* and *L. fabarum* are closely related parasitoids

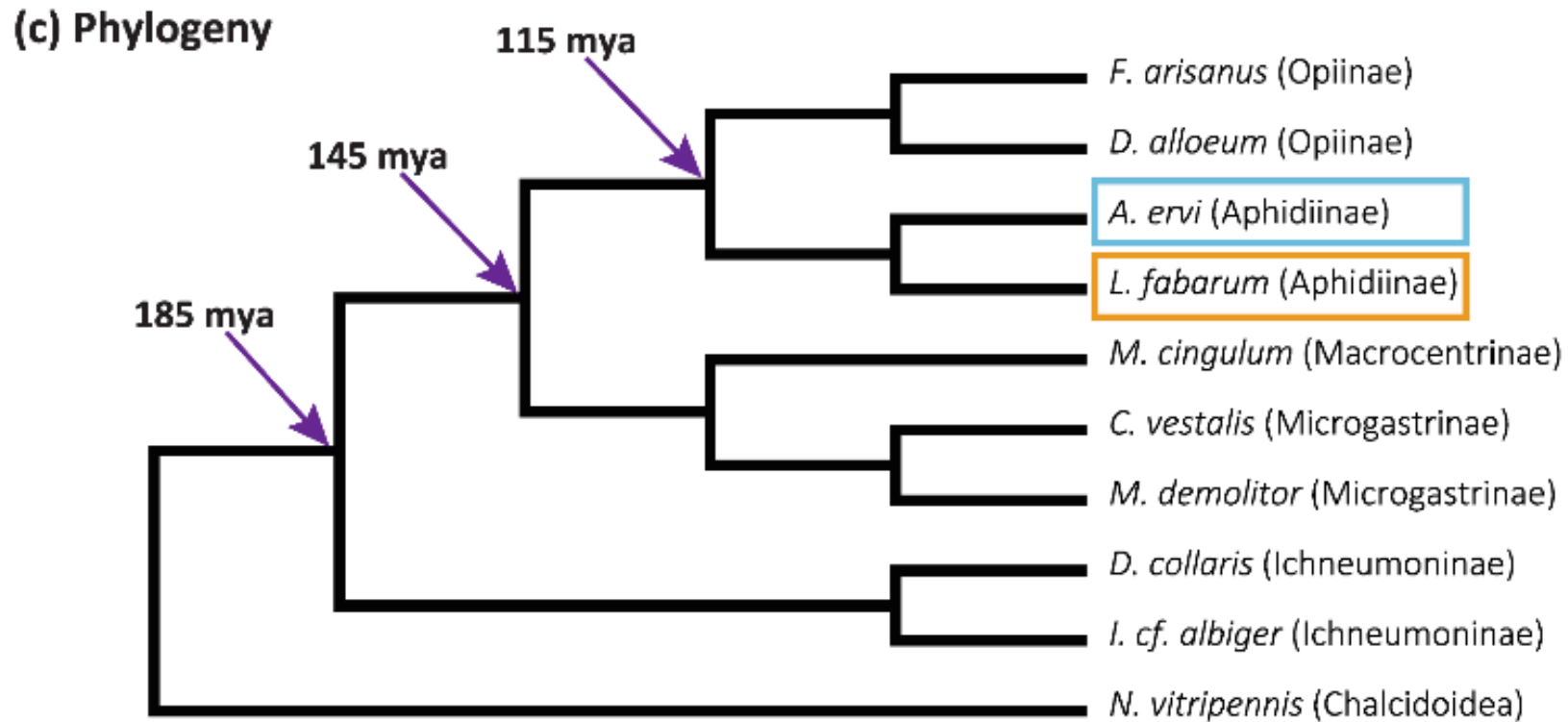
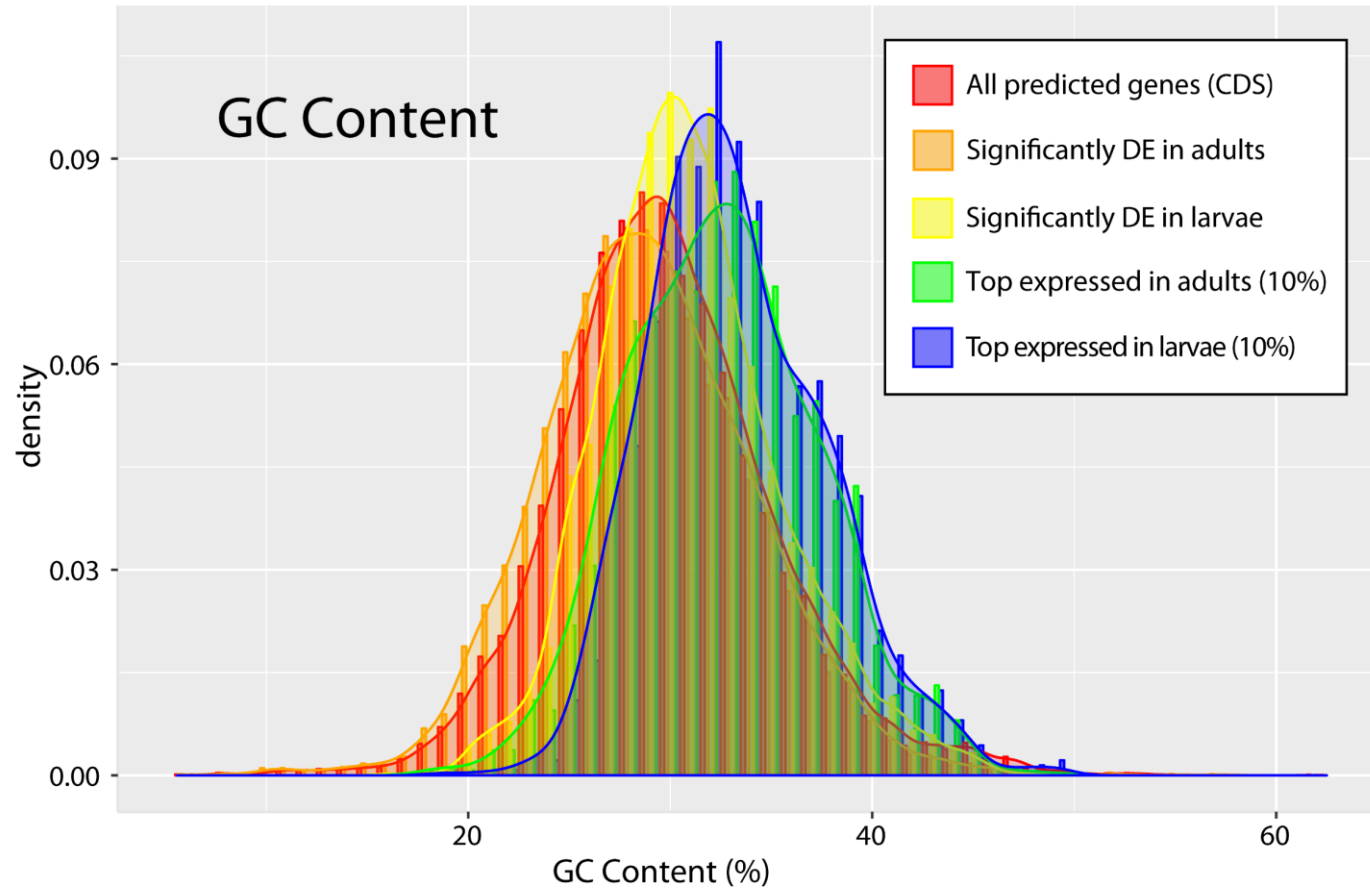
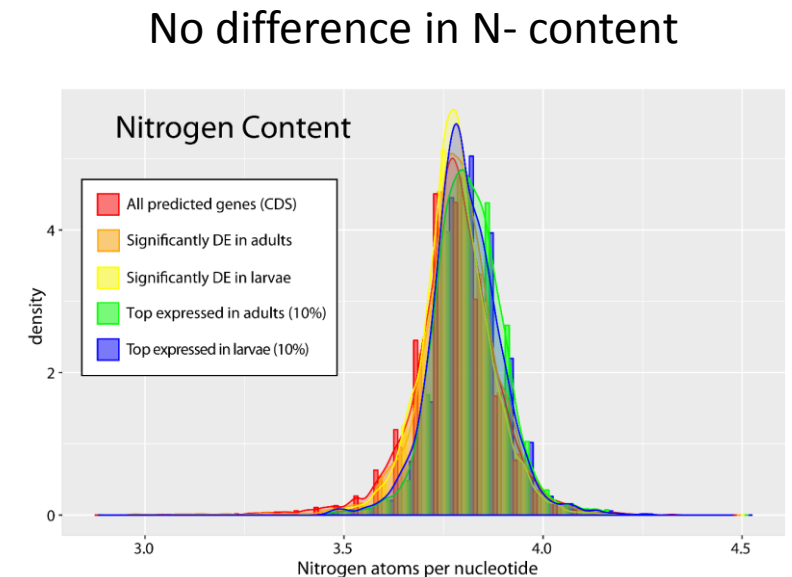


Fig. 1 Life history characteristics of two aphid parasitoids. Generalized life cycle of *Aphidius ervi* and *Lysiphlebus fabarum*, two parasitoid wasp species that infect aphid hosts. Figure by Alice Dennis. Life history characteristics of the two species. Phylogenetic relationships of the Ichneumonoidea species listed in Table 2, rooted with *Nasonia vitripennis* (Chalcidoidea). Average divergence times between major groups and phylogenetic relationships have been modified, after Supplemental Figure S19 in [1], *Ichneumon cf. albiger* is also included to better match dating available from [1]. The subfamily for each species is given after the species name

# Comparing GC of different subsets: *L. fabarum*

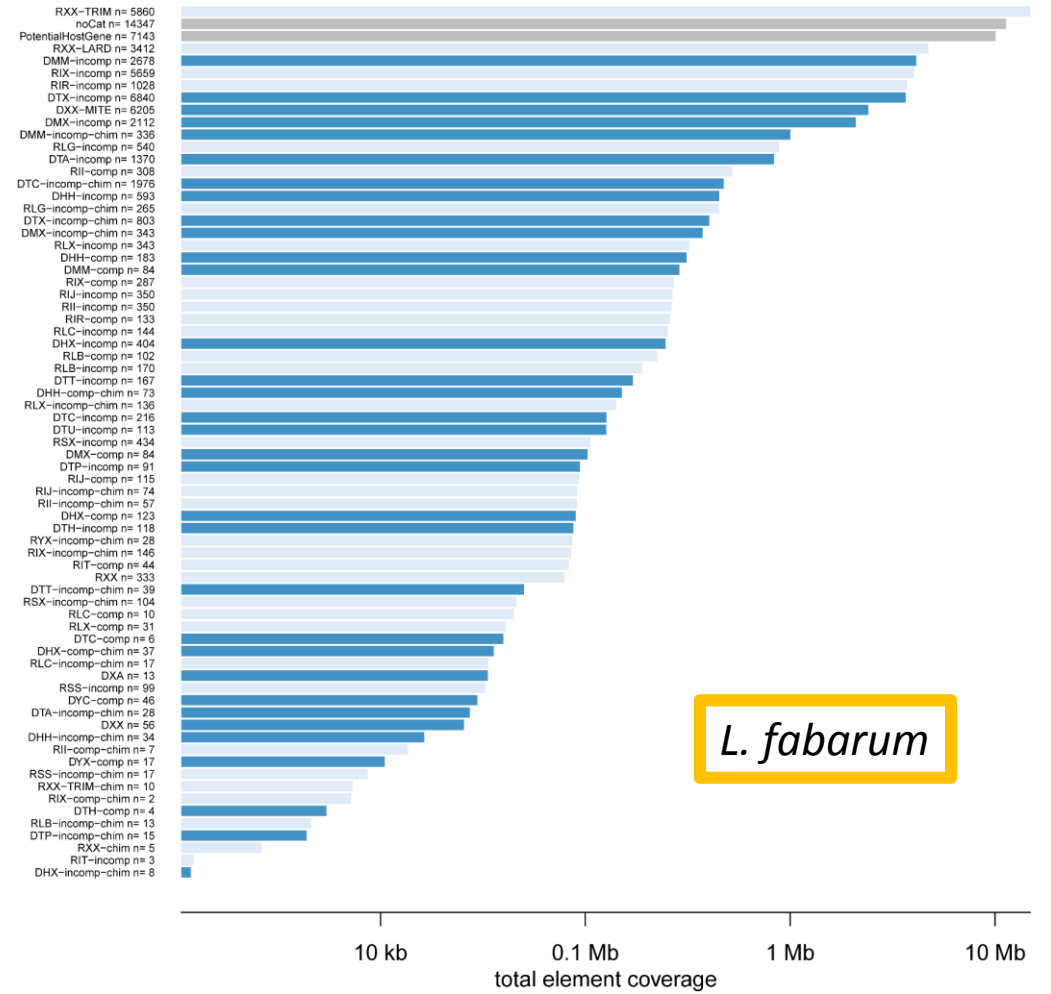
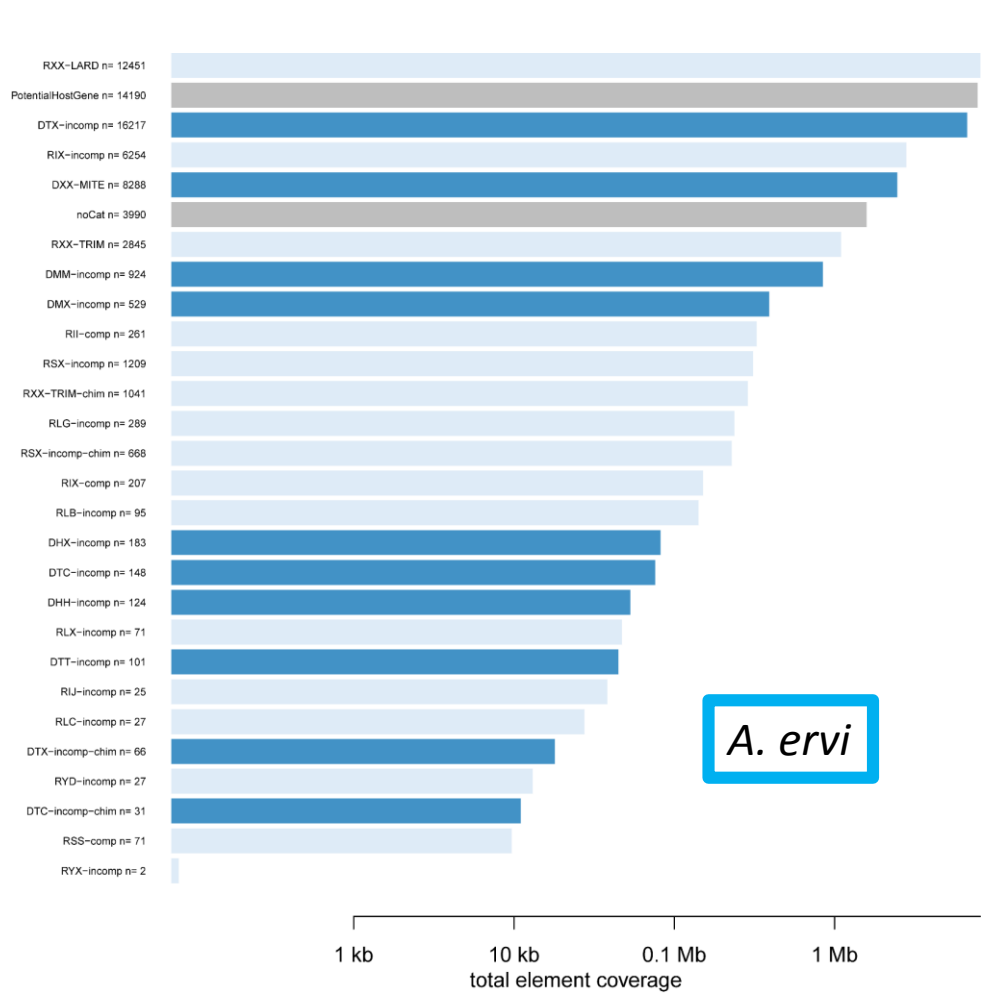


Lower GC content in adult-biased genes



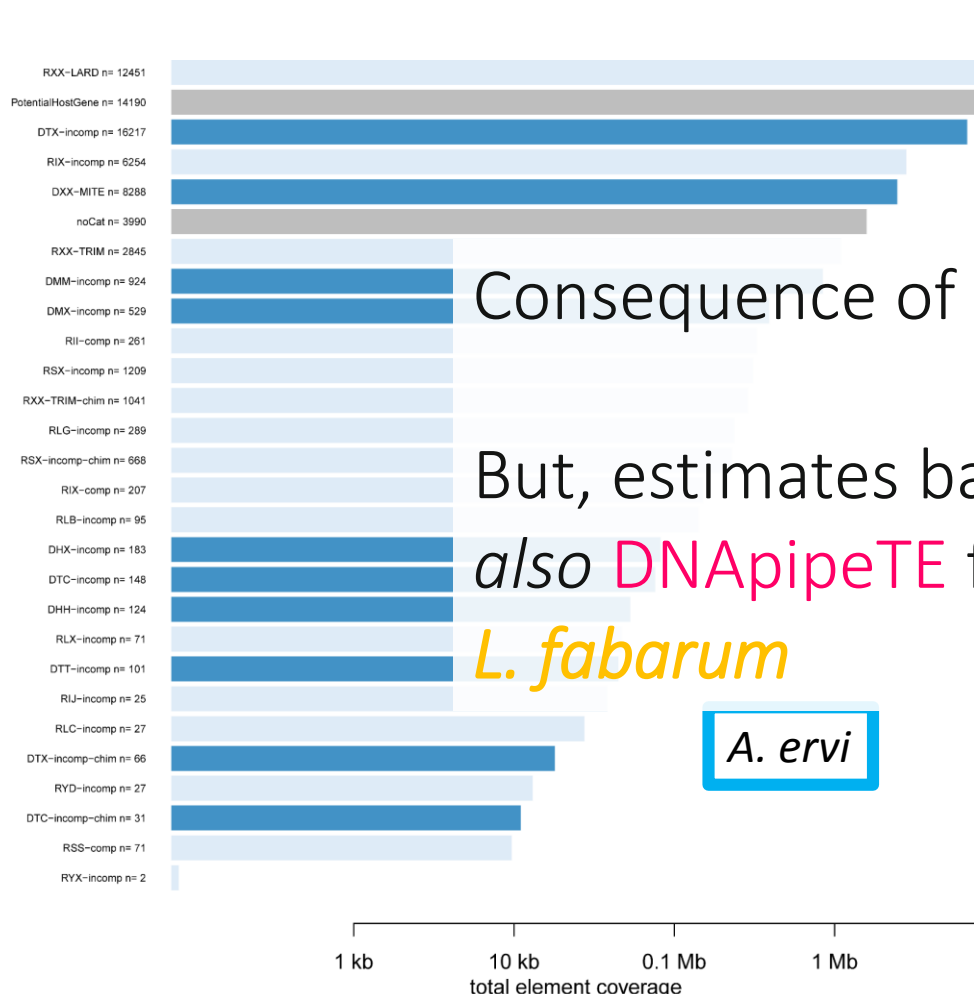
# Transposable Elements (TEs)

REPET (both homology- based and *de novo*)



# Transposable Elements (TEs)

REPET (both homology- based and *de novo*)



*A. ervi*

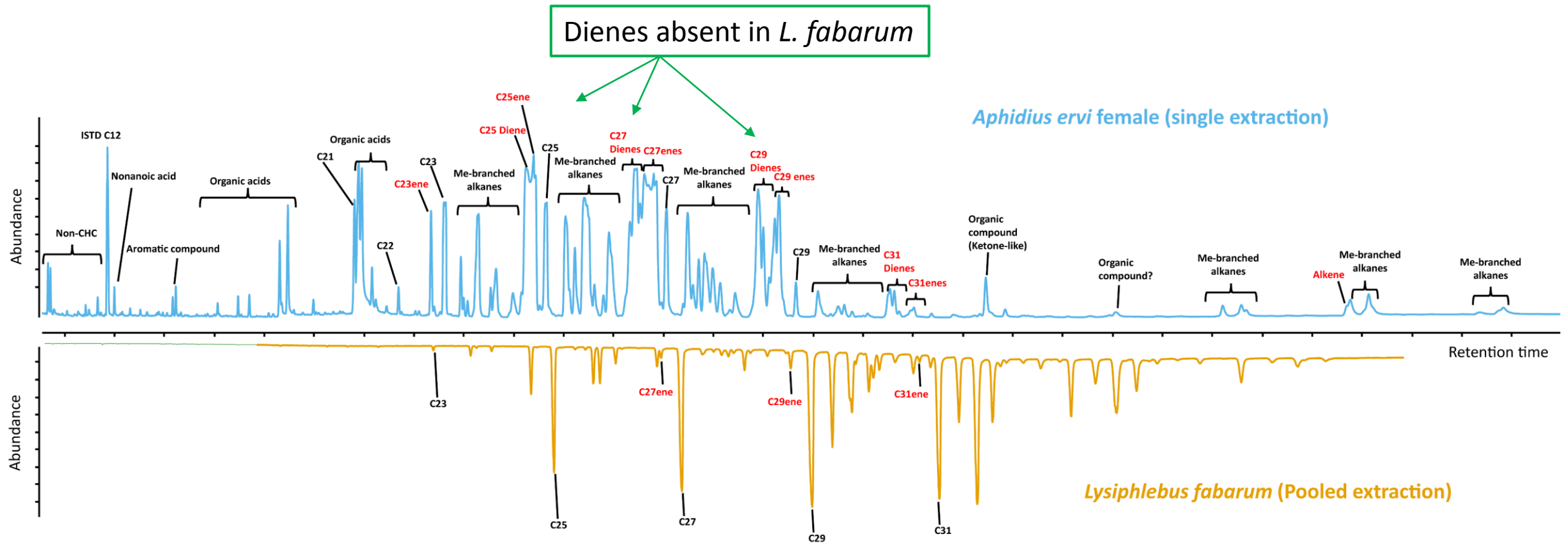


*L. fabarum*

Consequence of assembly? Maybe.

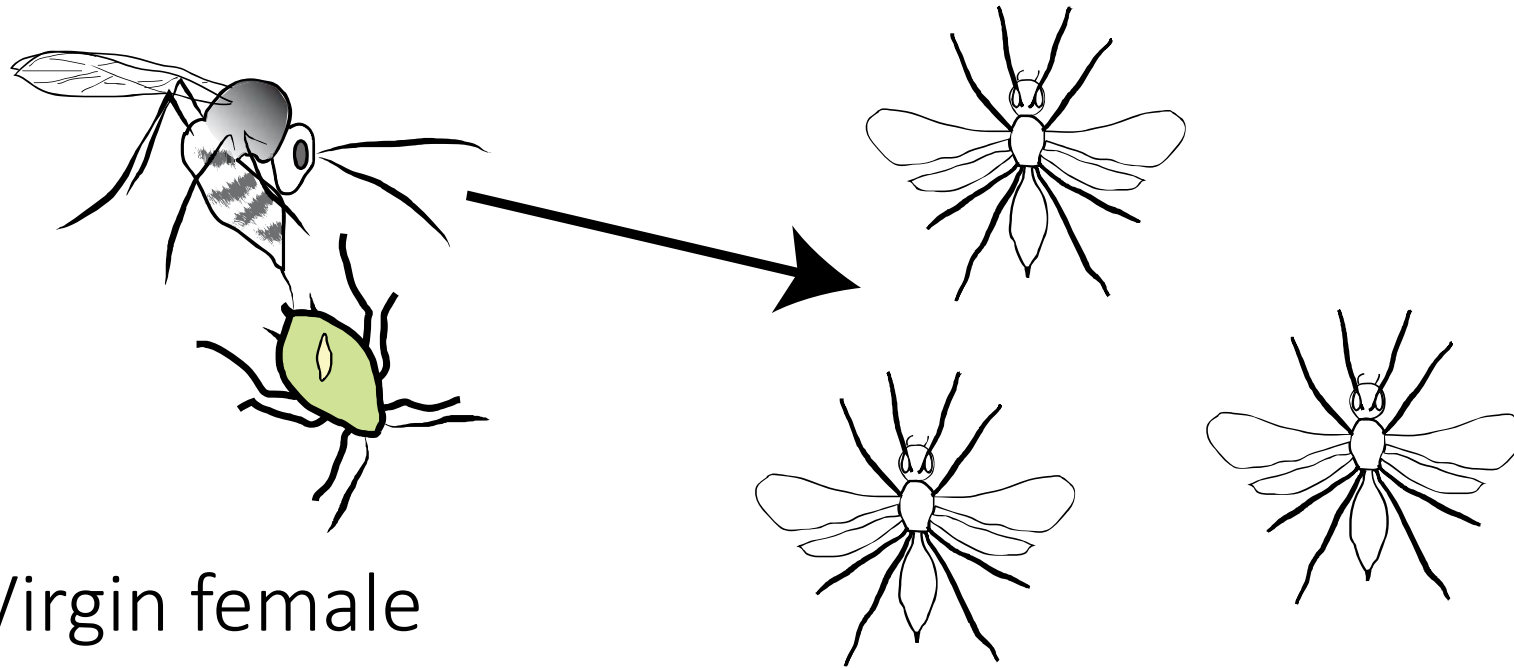
But, estimates based on the raw short-reads also DNApipeTE found higher repeat content in *L. fabarum*

# CHC profiles



# Linkage maps in *L. fabarum*

Taking advantage of haplodiploidy



Virgin female  
(sexually reproducing population)

Haploid sons

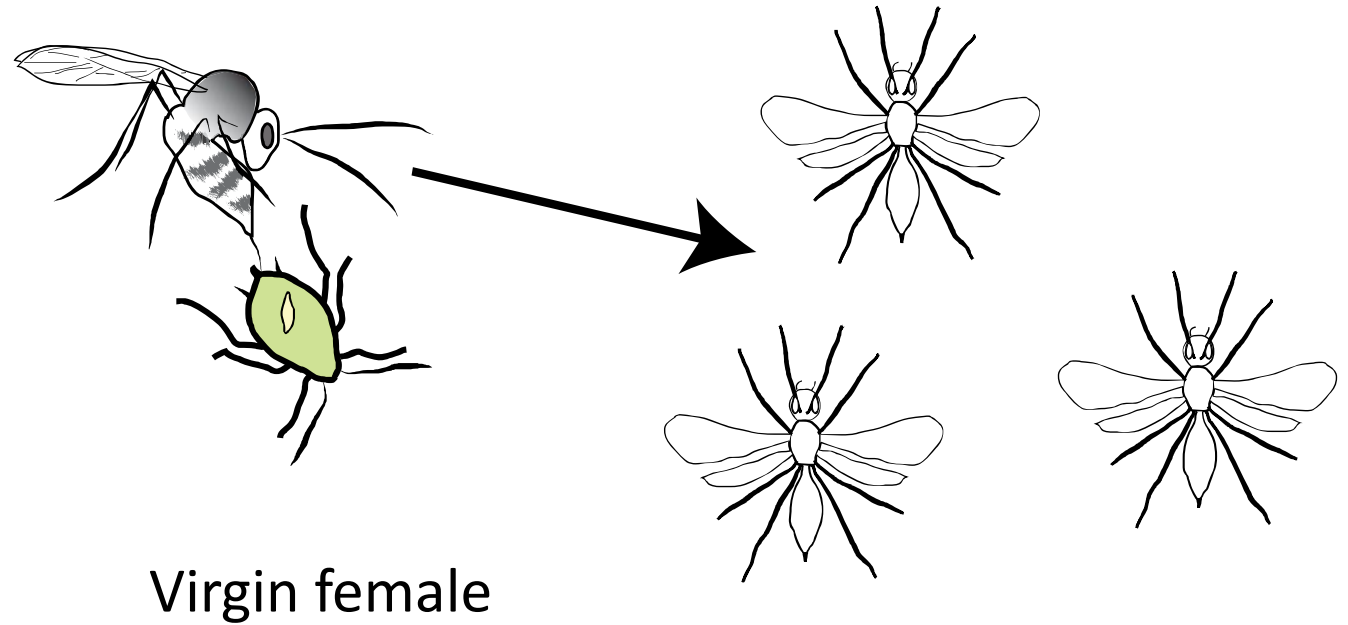


# Linkage maps in *L. fabarum*

ddRADseq of 124 males (**STACKS**)

filtering

1,319 SNPS in 90 individuals



Haploid sons

# Linkage maps

ddRADseq of 124 males (**STACKS**)

filtering

1,319 SNPS in 90 individuals

Mapping (**MSTmap**)

6 linkage groups

+



Predicted chromosomes for *Lysiphlebus fabarum*

