Ultraconserved non-coding DNA within insect phyla

Thomas Brody[1]* Amar Yavatkar[2], Alexander Kuzin[1] and Ward F. Odenwald[1]*

[1]Neural Cell-Fate Determinants Section,

[2]Division of Intramural Research, Information Technology Program,
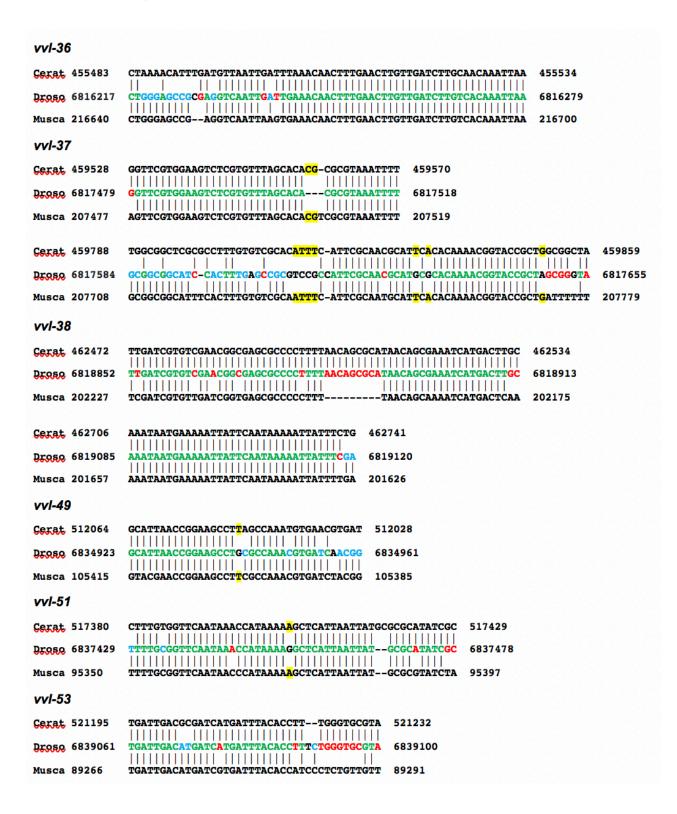*NINDS, NIH, Bethesda, Maryland, USA*

Running title: Ultraconserved sequences in insect enhancers

*Corresponding authors: Tel: 301-496-3435

 E-mail: brodyt@ninds.nih.gov,

Supplemental Table and Figures

Supplemental Table S1: Ant and Bee species formatted for *EvoPrint* analysis

Bee species (superfamily **Apoidea**)	Ant species (superfamily **Formicoidea**)
Apidae – social bees *Apis mellifera* - Western honey bee *Apis dorsata* - Giant honey bee *Apis florea* - Dwarf honey bee *Bombus terrestrialis* - Buff-tailed bumblebee *Bombus impatiens* - Common eastern bumblebee *Habropoda laboriosa* - Southeastern blueberry bee **Megachilidae** *Megachile* rotundata - alfalfa leafcutter bee – solitary	**Myrmicinae** – *Atta cephalotes* – Leafcutter ant *Acromyrmex echinatior* – Panama leaf-cutting ant *Wasmannia auropunctata* - Little fire ant *Pogonomyrmex barbatus* - Red harvester ant *Monomorium pharaonis* - Pharaoh ant *Solenopsis invicta* – Red fire ant *Vollenhovia emeryi* – Japanese ant *Cardiocondyla obscurior* – a tramp ant **Formicinae** *Lasius niger* - Black garden ant *Ooceraea biroi* - Army ant – colonial raider ant *Camponotus floridanus* - Florida carpenter ants **Ponerinae** *Harpegnathos saltator* - Jerdon's jumping ant **Dolichoderinae** *Linepithema humile* - Argentine Ant

Supplemental Figure S1

vvl-36

```
Cerat 455483   CTAAAACATTTGATGTTAATTGATTTAAACAACTTTGAACTTGTTGATCTTGCAACAAATTAA   455534
               ||   |       ||  |||||||| |||||||||||||||||||||||||||||  |||||||
Droso 6816217  CTGGGAGCCGCGAGGTCAATTGATTGAAACAACTTTGAACTTGTTGATCTTGTCACAAATTAA   6816279
               |||||||||||  |||||||| |||| ||||||||||||||||||||||||||||||||||||
Musca 216640   CTGGGAGCCG--AGGTCAATTAAGTGAAACAACTTTGAACTTGTTGATCTTGTCACAAATTAA   216700
```

vvl-37

```
Cerat 459528   GGTTCGTGGAAGTCTCGTGTTTAGCACACG-CGCGTAAATTTT   459570
               |||||||||||||||||||||||||||||| ||||||||||||
Droso 6817479  GGTTCGTGGAAGTCTCGTGTTTAGCACA---CGCGTAAATTTT   6817518
               |||||||||||||||||||||||||||||   ||||||||||||
Musca 207477   AGTTCGTGGAAGTCTCGTGTTTAGCACACGTCGCGTAAATTTT   207519
```

```
Cerat 459788   TGGCGGCTCGCGCCTTTGTGTCGCACATTTC-ATTCGCAACGCATTCACACAAAACGGTACCGCTGGCGGCTA   459859
               | |   |     | ||  |        |   | ||||||||||||| ||||||||||||||||||||  ||| ||
Droso 6817584  GCGGCGGCATC-CACTTTGAGCCGCGTCCGCCATTCGCAACGCATGCGCACAAAACGGTACCGCTAGCGGGTA   6817655
               |||||||||||  |||||| |||   |||   ||||||||||| |||| ||||||||||||||||||||||| |
Musca 207708   GCGGCGGCATTTCACTTTGTGTCGCAATTTC-ATTCGCAATGCATTCACACAAAACGGTACCGCTGATTTTTT   207779
```

vvl-38

```
Cerat 462472   TTGATCGTGTCGAACGGCGAGCGCCCCTTTTAACAGCGCATAACAGCGAAATCATGACTTGC   462534
               ||||||||| |||| |||| |||||||| ||||||||||||||||||||||||||||||| ||
Droso 6818852  TTGATCGTGTCGAACGGCGAGCGCCCCTTTTAACAGCGCATAACAGCGAAATCATGACTTGC   6818913
               | |||||||| || |||| |||||||||| |||      ||||||||||||||||||||| ||
Musca 202227   TCGATCGTGTTGATCGGTGAGCGCCCCCTTT---------TAACAGCAAAATCATGACTCAA   202175
```

```
Cerat 462706   AAATAATGAAAAATTATTCAATAAAAATTATTTCTG   462741
               |||||||||||||||||||||||||||||||||||
Droso 6819085  AAATAATGAAAAATTATTCAATAAAAATTATTTCGA   6819120
               ||||||||||||||||||||||||||||||||||| ||
Musca 201657   AAATAATGAAAAATTATTCAATAAAAATTATTTGA   201626
```

vvl-49

```
Cerat 512064   GCATTAACCGGAAGCCTTAGCCAAATGTGAACGTGAT   512028
               |||||||||||||||| |||  ||||| |||| |
Droso 6834923  GCATTAACCGGAAGCCTGCGCCAAACGTGATCAACGG   6834961
               |||||||||||||||| |||||||||||||| ||||
Musca 105415   GTACGAACCGGAAGCCTTCGCCAAACGTGATCTACGG   105385
```

vvl-51

```
Cerat 517380   CTTTGTGGTTCAATAAACCATAAAAAGCTCATTAATTATGCGCGCATATCGC   517429
               ||||  ||||||||||||| |||||||||||||||||||||  |||| ||||
Droso 6837429  TTTTGCGGTTCAATAAACCATAAAAGGCTCATTAATTAT--GCGCATATCGC   6837478
               ||||||||||||||||||| |||||||||||||||||||  |||| ||||
Musca 95350    TTTTGCGGTTCAATAACCCATAAAAAGCTCATTAATTAT--GCGCGTATCTA   95397
```

vvl-53

```
Cerat 521195   TGATTGACGCGATCATGATTTACACCTT--TGGGTGCGTA   521232
               ||||||||  |||||||||||||||||||   ||||||||||
Droso 6839061  TGATTGACATGATCATGATTTACACCTTTCTGGGTGCGTA   6839100
               |||||||||||| |||||||||||||| |        ||
Musca 89266    TGATTGACATGATCGTGATTTACACCATCCCTCTGTTGTT   89291
```

3

Supplemental Figure S2

A

tatttgagggccgcggagggaatttctttcaacgagggatcccgccgatcccccgcaaaaacaaagaaa
cacgcaaattgccgtcgcaaaaacacaacagaatatattgcattatatattaacaatataatttgttcatc
tcTAGTaacgTAGTATGTAGAAGCAGCTGTGATAAATAGTGCGCATATTATACGAAaaaaGCAAAAAgAGT
ATACGACAaACAAataaaaattaaaAgcAAccacAActactactataAAAataaccaaggaaaaggcaagg
aaactcgcaaaacgagagagagaggaaaagggaaaaaacgcCATGGGATTTATGACAACAACAAatgcaacc
cagtaacaaatgatgttgcagCAACtactacaatacacgagtataccaagaacatcaACAACaaCAActgC
AACGGCAACcAAcAACAAATTGTATACGATTTTTATAAACCAAACAAAtgcTATGTAAATAaTtaTtatAA
CAaacacacaagtaAAAAAaacAttcAAAAccAGATCTCCCGCTctagaaaacaaacAAcaAACGCGCGCG
TAACaaaaacgACAACAtCAACAatACAACAAGtAtACAacgtaatAATAATAcATAAAaTCTaAACaAA
TGTGCCTGTGTTAAAAaAAAAgAAagcaAAAtcAAGaagcaaagcgttcaaaaagcaaaggcttcccaata
catgaaaaaagtgcaaaacatatataaaaatccaagaaaataacatccaaaactgaagaactggaaaagcc
aagagtaaatctaaggactccacacgggcagctacagtagagcttccctaaaaccgaagcagttgcagcgg
cggagcgggacgtcggtcagtgtgttgcgcCCCcatttgttgttcctACtgtaAAAAGTGGCACAAATCGG
GTTAGACGTCGCTCCCAGAGCATTAATGGCTCAGCCcAGGGTAAGTAgatccaccaaagttcactttctaa
atatatagaatccatatagaatggtcccccgagagtgttgttatccttaaacaagttccgacggctaaaaa

B

Cerat 387653   GTTGAAGAAGCTGTGAT-ATTTGTGAGCATAT 387626
                 || |||| |||||||||| | | ||| ||||||
Droso 10613790   GTAGAAGCAGCTGTGATAAATAGTGCGCATAT 10613756

Droso 10613564   cATGGGATTTATGAC 10613577
                 |||||||||||||||
Musca 35636   CATGGGATTTATGAC 35649

Cerat 386904   TCAAAACCACATATACAAGCATTGTATACGATTTT----AAAGCAACAAAAT--TAT-TAAATTATT 386865
                 | | | | | ||||||||||||||||| ||| ||| |||| |||| ||| ||||| |||
Droso 10613458   CAACGGCAACcAAcAACAA-ATTGTATACGATTTTT-ATAAACCAAACAAAtgcTATGTAAATAaTt 10613522
                 |||| ||||| |||||||||||||||||| |||| ||| | || ||| ||||
Musca 35742   CAACAACAACCA-CAACAAAATTGTATACGATTTTTTCTAAAACAAAAATATAATAATTAA--AATT 35805

Cerat 386251   AAAAGTGGCACA-GGTCGGGCTGGACATCGCTCCCAGAGCATTAATGGCTCAACCTAGGGTAAGTACAT 386184
                 |||||||||||| ||||||| |||||| |||||||||||||||||||||||||| || |||||||| ||
Droso 10613042   AAAAGTGGCACA-AATCGGGTTAGACGTCGCTCCCAGAGCATTAATGGCTCAGCCcAGGGTAAGTAgat 10612975
                 |||||||||||| ||||||||| ||| |||||| ||||||||||||||||||| || ||||||||||
Musca 383950   AAAAAGTGGCACAAATCGGGTTGGACACCGCTCCTAGAGCATTAATGGCTCAACCTAGGGTAAGTATTG 384012

4

Supplemental Figure S3

Anopheles gambiae

Composite *e*BLAT

Score	Start	End
1420	1	1420
87	15	992
77	332	1061

Selected for *EvoPrinting*

Anopheles gambiae-S1

Composite *e*BLAT

Score	Start	End
1405	1	1420
84	916	1172
80	376	1118

● 1st ○ All Alignments

Anopheles melas

Composite *e*BLAT

Score	Start	End
1374	1	1420
73	322	517
50	332	484

● 1st ○ All Alignments

Anopheles merus

Composite *e*BLAT

Score	Start	End
1357	1	1420
76	332	1065
87	341	965

● 1st ○ All Alignments

Anopheles christyi

Composite *e*BLAT

Score	Start	End
940	50	1390
75	50	168
92	286	507

● 1st ○ All Alignments

Anopheles epiroticus

Composite *e*BLAT

Score	Start	End
733	107	1266
123	234	1414
84	329	716

● 1st ○ All Alignments

Anopheles culicifacies

Composite *e*BLAT

Score	Start	End
535	137	1382
105	476	953
86	332	429

● 1st ○ All Alignments

Anopheles funestus

Composite *e*BLAT

Score	Start	End
534	137	1384
131	285	974
99	33	356

● 1st ○ All Alignments

Anopheles dirus

Composite *e*BLAT

Score	Start	End
447	145	1265
86	522	1028
101	2	903

● 1st ○ All Alignments

Anopheles farauti

Composite *e*BLAT

Score	Start	End
413	145	1172
85	7	531
84	5	300

● 1st ○ All Alignments

Culex pipiens

Composite *e*BLAT

Score	Start	End
204	44	1172
109	31	578
85	44	922

● 1st ○ All Alignments

Aedes aegypti

Composite *e*BLAT

Score	Start	End
181	245	824
172	245	824
71	761	1055

● 1st ○ All Alignments

A

Apis_Mellifera (Bee) (Ref Sequence)
Composite *e*BLAT

Score	Start	End
2808	1	2808
322	717	1780
227	438	1652

Selected for EvoPrinting

Apis_dorsata (Bee)
Composite *e*BLAT

Score	Start	End
2388	76	2788
284	700	1654
245	1223	1796

● 1st ○ All Alignments ○ None

Apis_Florea (Bee)
Composite *e*BLAT

Score	Start	End
2299	61	2808
253	1143	1826
195	968	1853

● 1st ○ All Alignments ○ None

Bombus_Terrestrialis (Bee)
Composite *e*BLAT

Score	Start	End
917	261	2497
127	881	1800
122	1305	1985

● 1st ○ All Alignments ○ None

Bombus_Impatiens (Bee)
Composite *e*BLAT

Score	Start	End
860	261	2462
196	481	1806
159	677	1811

● 1st ○ All Alignments ○ None

Habropoda_Laboriosa (Bee)
Composite *e*BLAT

Score	Start	End
762	287	2453
259	681	1810
142	1140	1947

● 1st ○ All Alignments ○ None

Megachile_Rotundata (Bee)
Composite *e*BLAT

Score	Start	End
307	2070	2450
119	1589	1798
127	1705	1894

● 1st ○ All Alignments ○ None

B

```
aaaacgttatatatttattatatgtataatatttgaatagtgttttgtgacagtcatgacaatcattgacta
ttatatatataactcaaatcatcaagttatattgagagttcaagacaactcgttgataatatatatccaa
tactatcgtagcgatatattttcgataatcgttaataattcattgacaaagttgtgaagcaataattttta
caaaagttttttgaaattatctaaaatcgattgatctctggaacaccctatatatcttgttgtatgttctcg
TTAaATctatccCtatataTACATCTATCAGGATCAGCCcGCAATTAATTCcacacgCAtAtTTCATCCgtg
cgaatGTacacatttACACaaggattaccgccatgtccactgtctcgaattcttcgacaagcaacgattttt
tccaaagaggattggcgaacatttatacgCtCGAAATTGTCGCGaGAAAcGAGCGAgAAGTTTTAATTGCtc
cGCgtACgAGCgCaCCaATtGAATCgTCTTtGTgcgACGATATTGTTTaAatacACaccttcgAAAGGTTtc
aTGCTTACaACTCGTTAAAagtgcagacatgtgcagaGAcaAAACggtgtatcgaTACAATcCGGTTCccCG
TTTCATTGTtccgaacatacaaacacattgttagatttctttcctttttttttattttttttttcatatttt
TCTttcTtTcTctttttttttagattttttttttTGTTTGcaATTTAATtCaACaTTCGAAtgGAAAATtttac
tacggtGAacagagcgGAAAAAGTTTaTAAATGGAGcGCgtAaATTTAAcgttCgcCgATTAAAtGAAAGTG
ttCGTTTggtcggttTTtttCAGaGgAATCGACaGgAAATTGtatGaATTAATcGCTCGaATTcgAAtgaCa
aATAATgtaaaaggattcatgtaatgtttttattaaatttaatttaataaaaaaaaggaaagacaagata
tttcagtttgcaacagaatgaatcattcaaactgaaatagaaatattttacaatatcgtttgttttgact
atttttgtttatttctctttttttgtaattttttcttttatttacttttttaaagtcatatttaaatgt
ttttgaaaaatatttacgtatttttatttttatttcactataaagttattaataatgatattgtaaaa
ttttatataatttatcatgtaaaacattagtattttaaattgcaattataacatatcaaaatatgtaaagta
tataaaagtttattagtttacatattaatattatgcaaattttatctatttattaagttagttttcattc
atctttccaatataattgtaaaaattatttatctttcttttactatttaacatcatgcaacaaataatttca
aatatttcttatcattataattctaaaaattatttaatgatcttttctttctaacaattttagaaaagaaa
ttatttatacaaatatttataacttttattcttcaaaacaaataatttatgtcaaacttattcaaaagtat
cattttctcatttattaaatatatgtttgtttaaaacagttaacatttaaaatattgttttttcaatatt
atcaacaaattcaaagatattggaataataattaaaaaatattcgaaaaataattaaatagtctgtatatac
atatttcaatttacttataaatatatatatacatactcttactatatatacatacttatatatttatttataa
atatgatctatttcgatcactattctaaaaaaattaaagaaaaataataaaaaaaaaaggaaaacaacaatc
taatttagaaatgaacaacaatctaacacaaacaattgaatacaaaaactcatattctttcctttcttttta
ttatttaatgaaaacgcatacacaaatatatttacaaattctactcacgtaatcaaccttgaaccgacgcat
tcgtacatgatataaatttcacaggcgcgtctcaTATTTCTtGCgGCctAaATAATATTAcaaCGCCTTTGT
TCcCCttcagtgagtaCTATATTTCcGTAAGCGTAACATATCCATCTTACTAATAGCGCtATGGTCAGACTG
CACGTAATTCGAGGtAGTCaccaacggttccccctgtggtatatatctcctgtATGTaCAtgtcgtatatgg
ccccgtaTACATACGcgGtaTATCcAtgTTGAATaGATCCGAacAGGCAgGGcATAGGTAGcAGCATCGAGT
GTtgATACCGAAACTACATAACCtATCCAAGCCGGCTGACTTGGCCGACAATCGCTATTTAGgaagCCGCTTA
TCTCaGCCaGTCTTAGCTGatCTAATGTCTCttggCGAATtTGcTTtAaCacGTAGAtGcAATAACGTatac
ggGAaTcttagatagaaatgttgtgtttagaacgaagcaagaaaaaggaaaactaaaagcgagaagggaaaga
ggaaaaaaaattgaagggataaggaggtaaaagaaaatggcaaaggtgaagagaaagtcgtgaggggaatg
gggtgagagtggtcgtgaaaatggaagaaatttggagcaaggatgataatacgaaaagagaaaagtgagaa
ataggatagacaagaggaggagtgagtaatatcatagaaagaataaagaatcaaggaaaggattaaaaaag
atacaatttattggagagaaaaatatgaaaggattattgttaaaatatgtatagtttttatgtcaatttta tt
```

Supplemental Figure S5

A

tggaatatgttgaagcaacgacggaatatttcgtatataattttgattcaaaattgaaatcgattgattaatttattataa
aatgaatgtctttgtcgaatgctttccttcatcttcaggaaatccttggttaattcggaaataaatagcttgccccgcggt
tccgcgtatattacacgcgtccctcccccccgcggcgtatggatctcgtgGCGtAAAAATGAAAGAAGGCGcAAATCCTCG
cgggtggaccgaCGCgTTCcGCAAcAAAATTTCCTTtATGCCCcCgAgGGCTTTAAGTATTTCACCCGCAAGAACCtcggC
GCAGATTGGGGaTGCGCaggCTGAAATCCGCAATCCGTTTCGATCACGCgCGAACGATTTAGCTcCTTGCAATCCCGtgtt
GGAAGAtgaAAAACGAGAAGGAAAAAATATGCAACggcatcggcggttctcaatggagagagagagaggagcactaggggca
cggaaagagaaagagggagagaggggagagagagaGAGAAgcgcgagtgacgccgcgcGTTGAAAGCATAAgcatccaCCC
TTCGCACATTTTTATGCcggAGACGTCGCCTCTAAGCAGCTTTCTTCaAGATTTTAAACCAAAACATCCCTTGGGacccggg
gatggttcccagtatttttcccacttcccttttcacgcggaataaggttgcgaattgttgagaaacgttattacaattt
cgtcgctctatcggcgataaacgcgaaacctttcacgtttcgattaatcttctttttaactcctccttggtcgtttcctttc
cttccatttttaagaaaatagtaacgctttttaaagtgatctaacaggcatttgttttactcttatttgggaggctaagaa
tgacgtttaatgaaaaatccataagaaaagaatatttttttcatatctttttacgaaattctctaaatacgtaaaatatat
aacatacttccttatgattctcttcctttcgtttcatcgatgtcttcgatcgaaacgaaatgagaaatgatatctgataa
gaatatagttcaactattcattatttccagaaatggaacgattattatcaaagagatggctcttttgataaaaatttttcg
tttcatcgaagtaaagattttaagaaatgggattaaTTTAAtattattaacgatagttTTtTcTcgttTTCAGCtGCAGCA
TAtCCtAgTTTtTAtTCCGGcGGtTACGTgGCaGcAATGGCAGCcGCCGGCGTTTCgCTtCACGGtCAaCCGGCtTTCTAt
CACGCgGCTCATCCcTAtCAAATAAATCCAAAcGCcatGGCtgtgggGCCgATGGCgAAgAGGAAGCGTCGtCAcAGGACG
ATATTcACcGAgGAgCAgctGGAACAAcTgGAAGCcACtTTCGACAAGACcCACtTAtCCaGAcGTaATtttTGAGAGAACAg
CTcGCtCTtCAgGTCGATCTCAAGGAgGAacGAATCGAGGTGAGTttagacaaaataagttgtatgtttcaagatcttacg
acacgtgtattgcagtacatgattatgaaatttcagttttttgaatgtaagatttttctttttgaagaaatattccccact
agtgtttgtaatttggatttattttttgttcttcgtacttatattgatagtttttctattttttcaattaaatttcgttaaat

B

tgacaccggcgcttattattaaatcggccggagttactcgacttcgcgctggagaggcgatccgattcgcgtgagtgccga
gagaaaactttcgcggtttcgcgtagaaataatgtgaggcaaaaaggcactgtcttatccaaagatgtatatttcttcgtt
gtagcgagaGGGAGAATAATacgcgccgaaTTCTCCCttgccacaCtTTTTTTccctctcggtaacactccgttcctctt
ttTcgcTTtCTCTCTTTCTTcCTTTCTTAGCctctctcccttctctttctctctctctctctctctctctctcttTCtC
TctttctccctgaagactaAActCTTTCTCTTTTTgcaccgggatgcgtATACGtTCTtGCAGAAaCgctgcACGGGCcGc
CAAtGTACGCggcgttccctccGtTcCTtACGGCACATATACCACgaaGAGCgGGGAGTCCCCGcGAGCCaCGAGAAATAT
TCGACAATTTTTAAACTCGTTCCTGAcAcacgcgGGCATTCAAACgccGCGCaccagcctctcgcagtcagCAAGTAcGAAT
TTCTGAGAAtttcggccgCGTCCTGGTaGGGtGcACcgAAtgctggatgacgcgcgtgatagatgttggttgggattagaa
cactcgagcccgaggaggacactccgggtttcgaccgatgtgacacgtgatttgttgtcgttctgtgatcggccccgcgct
gcgtccgattaatcaacgagaagcgcagtcgcgatgcgcctttgttcgtaattcgctgatatggcccgactaatacatata
cgattatcctttcatgcatcaaatctctatagccactcaattgtgtagatttgcatttcttgggtgtacaatgtatcaatc
tagtgcccggcgggaaagtacgagtataggacggctcgtggaagagaggtttgccgcgtccaccggtattcgacatgcccg
AATAATAAAaTTTAAATACaTcctGGCAAGAAATGcAAATGcGGacCGCTtGAGAGtGCTTTAACTGTaGAATtatTTCGTTt
AgGCAAAAAAAaagGGgggagaatacggcatggtgtatgtgtgcgtgtgtgtgcgtgCgCAGtTtaTTTCCATCTCTCGTG
tGCGCATTACCTCGGGAGATAATTATCGCaAGCGAGGATTGGTCAACCGAATTGCGTCTcGCGCTCAATATTTTCTTGCGT
GcCaCGGcCGTttCgAGtCTTTTAGtGAGaGCtATGAACGCcCggagagtgaaatcgcttctaCGCGCCaccctctccaacc
tccgcatcGCCCGCGGCCACgCCaCGCTTTTTGCATAttGCGTATTGCGATCGGcCCcGATAGTTctttcttttttttaag
cCgACCaTTCTTtccccgcTCCccgCgcgCTctccTTTTTTtcctccgtCCtCTctccTTTATtTCGtGGCgcgtatccgtgc
gcgcgcccatcGgCAACCAGTTTTCGGtcgcgcgcgcgGGAGAGtgCCaTAGGTCGACGAACGgtgcGcTCaTAACTCgC
GgTATGTGTATCAGaagtcattccggtatgcgccgttTCtTCTGgcgatGTGAACGCgaacccgtattatccagaccgata
gtatcgcccagaacACcCGGTCGGTAAAataAtAGAtGGTTgcCTgtgGAgGgacgtgtttatgccagaccgagccgagttca
tatcgaccgaatacccatctggctcgcacggaaagagacgaagacctcgcgcagatttctagggcagtttgtgatctactg
agatctgtcgagctttgaatttgaaatttgattttgatagcaactttttagacatttatagtatttcacaaatatttcactt
attataacacgttcattgtacattaaatcaaaattatattataacgtaatctactttgaataagaaacggagaattgttgt
cgaaatgttcttgcttatttgatttgattttttgtgtaacattttcttcttgtatattttacaaatatttcacctattata
ttatacactagatcaaaattatattaacgtaatatctactttaagttagaaacggagaattgtcgttgaaatattcttgct
tatttaatttttattttttcgtgtaatatttttttgtacgttgaattaatttaaatatgtagtcgattcaatttaatatataa
taagagtgtttctcgtctgtcaCAGCcTTGtTCAGGCAGAGtATGTCtCCGGTaGCcGGAttCTcTgATGaGtGGTTCCAAa
acgtcgGTGTCcCAaCCgATCATcGATTAtTCcCGATAcGTGAGgaGAtTataCgTCCagcCAaGAATGCGGcAGTACGTAc
TGCAAgGAacTgGGtTGcAGgGAcCAcTTtCAtTGTcTtGACtTGtAGCGGcagggtgTTcGTGAAgAAggagGAaATGATC
AGGCACTTCAAGTGGCAtAAGAAgAGgGACGAgTCcctGCAGCAcGGcTTCATGAGgTAcTCGCCGAcGGACGAtTGtTCG
GAgAAgCAcCCcGGcCGgaCcTGTCCGCAcAAtaGAAAGCaGACcCAcTAcCAtTGCATcCACGAGAATTGCGACAAcGTg
TAtATatTCgACaTCGGACGTtCAgATGCACGCaAATTAcCAtCGcAAgGACTCgGCcATCATACAaGAGGGcTTCCAGCGA
TTTCGgGCcCACcGAgAGTTGCGCGACtGAgcATTGTccgTTcGctGGaCAaCGgACcACGCACTTTCACTGTCGtCGAcCg
GgaTGTCGGTTCACtTTCAAAAATAAtGCCGATATGGGTgagAAgcgaactgtttgcgTGATtAtTGcTCGCtctagatcg
agactattcatcAATCgaagCCTTtcactagaaacttaatgtgtcgcaaacgtttcggatttaattcttacgctgagagaa
cgattacgttgatacatttttatttgatgtacatatctaattatattttgtcgttatttcagatctctccatgatttgctc
tacaagttttcttttttattttttttttttcagacaagcacaagtcttatcacatcaaggacgaacaactgtcgcgcgacgg

Supplemental Figure S6

Apis_Mellifera (Bee) *hth* **Genomic Relaxed** *EvoPrint*

Cerapachys_Biroi (Ant) *hth* **Genomic Relaxed** *EvoPrint*

Supplemental Figure 1. Ultra-conserved DNA in *Drosophila vvl* enhancers identified in *Ceratitis capitata* and *Musca domestica* orthologous DNAs.

Three-way *Ceratitis-Drosophila-Musca* BLASTn alignments of CSBs within six different *in vivo* tested *Drosophila vvl* enhancers. *Drosophila* sequences that are shared with *Ceratitis* and *Musca* are shown in green. Red bases are shared only between *Drosophila* and *Ceratitis* and blue text represent bases shared exclusively between *Drosophila* and *Musca*. Yellow highlighted *Ceratitis* and *Musca* bases are not shared in *Drosophila*. Flanking BLASTn designator numbers indicate genomic base positions.

Supplemental Figure 2. Ultra-conserved *Drosophila*, *Ceratitis capitata* and *Musca domestica* sequences within the *homothorax locus*.

A) A 1,065bp *EvoPrint* of the *D. melanogaster homothorax locus* that includes 5' non-transcribed sequence, its 5' UTR, the first five codons of its encoded protein and 102bp of the first intron. Capital letters represent bases in the *D. melanogaster* reference sequence that are conserved in *D. simulans, D. sechellia, D. yakuba, D. erecta, D. ananassae, D. persimilis, D. grimshawi, D. mojavensis and D. virilis* orthologous DNAs. Lower case grey letters represent bases that are not conserved in one or more than one of the test species. *Drosophila* CSBs that are shared with *Ceratitis* and *Musca*, as detected in figure 1, are shown in green. Red bases are shared only between *Drosophila* and *Ceratitis* and blue text represent bases shared exclusively between *Drosophlia* and *Musca*. The translation start codon is marked by an underlined ATG. **B**) BLASTn two and three-way alignments of the ultra-conserved CSBs. Font color annotations are as in panel A. Yellow highlighted bases in *Ceratitis* and *Musca* are not shared in *Drosophila*. Flanking BLASTn designator numbers indicate genome base positions.

Supplemental Figure 3. Combined Ant and Bee *EvoPrints* identify ultra-conserved *Hymenoptera* DNA

A) An *Apis mellifera goosecoid (gsc) EvoPrint* generated with four evolutionary divergent bee genomes and then overlaid with a print that includes the four bee genomes plus four divergent ant species. The *Apis* honeybee *gsc* DNA (1,701 bp) includes 5' non-coding, the first exon and intron sequences. All uppercase bases (both black and red font) are conserved in bees and sequences that are conserved in both bees and ants are denoted with red-font uppercase bases. Lowercase gray-

colored bases are not conserved in one or more of the bee test genomes. Bee test genomes: *Bombus terrestrialis, Bombus impatiens, Habropoda laboriosa* and *Megachile rotundata*. Ant test genomes: *Linepithema humile, Monomorium pharaonis, Wasmannia auropunctata* and *Atta cephalotes*. **B**) *EvoPrints* of the ant *Wasmannia auropunctata castor* (*cas*) gene locus. The 3,078 bp *Wasmannia* genomic DNA includes *cas* 5' non-coding, the first exon and flanking intron genomic sequences. The initial *Evoprint* was generated with four evolutionary divergent ants and then super-imposed with a print that included these four ants plus four bee genomes.   All uppercase bases (both black and red font) are conserved in the ants *Cerapachys biroi, Linepithema humile, Atta cephalotes and Vollenhovia emeryi*. Sequences that is conserved in both ants and bees (*Apis florea, Bombus impatiens, Habropoda laboriosa and Megachile rotundata*) are shown as red colored uppercase bases. Lowercase gray-colored bases are not conserved in one or more of the ant test species. The translation initiation codon is underlined. The left flanking vertical brown bar indicates an ant-specific conserved DNA cluster that is not found in bees. Note, in the exon ORF most, but not all, of the conserved codons do not have conserved wobble positions indicating that the cumulative evolutionary divergence of the test species used to generate the *EvoPrint* afford near base pair resolution of essential DNA.

Supplemental Figure 4. Conserved sequence clusters within the honeybee *dscam2* gene

EvoPrinter analysis reveals *Apis mellifera* non-coding sequence elements that are conserved in other bee species or only in a subset of species. **A)** Alignment data generated from one-on-one comparisons of a 2.8 kb sequence from the honeybee 16 kb *dscam2* second intron. For each species, the top three independent *e*BLAT alignment scores are listed. Scores indicate the total number of bases within the reference sequence, the *Apis mellifera dscam2* intron, that align with the test species genome. The test species; *Apis dorsata, Apis florea, Bombus terrestrialis, Bombus impatiens, Habropoda laboriosa* and *Megachile rotundata* are listed (L -> R) based on their highest alignment score in descending order. Website links to individual *e*BLAT alignments and superimposed composite *e*BLATs are indicated in either red or blue font colors. As indicated in the alignment scorecard by the blue selection buttons, the top (highest scoring alignment) for each test species has been selected for *EvoPrinting*. **B)** A color-coded relaxed *EvoPrint* of the 2.8 kb honeybee *dscam2* second intron generated

from the alignment data shown in panel A. Black uppercase letters indicate bases conserved in all test species. Font colors represent sequences conserved in all species except for *Apis dorsata*, *Apis florea*, *Bombus terrestrialis*, *Bombus impatiens*, *Habropoda laboriosa* or *Megachile rotundata*. Gray lowercase nucleotides are not conserved in at least two of the test species.

Supplemental Figure 5. Combined Ant and Bee *EvoPrints* identify ultra-conserved *Hymenoptera* DNA

A) An *Apis mellifera goosecoid (gsc) EvoPrint* generated with four evolutionary divergent bee genomes and then overlaid with a print that includes the four bee genomes overlaid with an *EvoPrint* that includes four divergent ant species. The *Apis* honeybee *gsc* DNA (1,701 bp) includes 5' non-coding, the first exon and intron sequences. All uppercase bases (both black and red font) are conserved in bees and sequences that are conserved in both bees and ants are denoted with red-font uppercase bases. Lowercase gray-colored bases are not conserved in one or more of the bee test genomes. Bee test genomes: *Bombus terrestrialis, Bombus impatiens, Habropoda laboriosa* and *Megachile rotundata*. Ant test genomes: *Linepithema humile, Monomorium pharaonis, Wasmannia auropunctata and Atta cephalotes*. **B)** *EvoPrints* of the ant *Wasmannia auropunctata castor* (*cas*) gene locus. The 3,078 bp *Wasmannia* genomic DNA includes *cas* 5' non-coding, the first exon and flanking intron genomic sequences. The initial *Evoprint* was generated with four evolutionary divergent ants and then super-imposed with a print that included these four ants plus four bee genomes.   All uppercase bases (both black and red font) are conserved in the ants *Cerapachys biroi, Linepithema humile, Atta cephalotes and Vollenhovia emeryi*. Sequences that is conserved in both ants and bees (*Apis florea, Bombus impatiens, Habropoda laboriosa and Megachile rotundata*) are shown as red colored uppercase bases. Lowercase gray-colored bases are not conserved in one or more of the ant test species. The translation initiation codon is underlined. The left flanking vertical brown bar indicates an ant-specific conserved DNA cluster that is not found in bees. Note, in the exon ORF most, but not all, of the conserved codons do not have conserved wobble positions indicating that the cumulative evolutionary divergence of the test species used to generate the *EvoPrint* afford nearu base pair resolution of essential DNA.

Supplemental Figure 6. Side-by-side comparison of conserved sequences within ant and bee *homothorax* loci identifies shared exon/intron architecture and species-specific conserved sequences.

EvoPrints of bee and ant genomic DNA that includes *homothorax* (*hth)* encoding an exon isologous to the 2nd exon of *Drosophila hth* plus flanking intronic sequences. Blue and red underlined regions are coding and non-coding sequences, respectively, and align with homologous regions in the two panels. Black uppercase bases are conserved in all test species and colored uppercase bases are conserved in all but one of four bee tests species in panel A and all but one of three ant test species in panel B. **A)** Relaxed *EvoPrint* of *Apis mellifera* genomic sequences (6.3kb; Group5:7,111,526-7,117,900)*.* Vertical red bars flanking the *EvoPrint* indicate conserved bee-specific sequences that are not found in ants. Colored uppercase bases are conserved in all but one of the color-coded test species: *Apis florea, Habropoda laboriosa, Bombus terrestrialis* and *Bombus impatiens*.   **B)** Relaxed *EvoPrint* of *Cerapachys biroi* genomic DNA (5.1kb; 6532628-6527517, *Ooceraea biroi* isolate clonal line C1 chromosome 14, Obir_v5.4). The green vertical bar flanking the *EvoPrint* indicates ant-specific conserved sequence that in absent in bees. Black uppercase bases are conserved in all test ant species and colored uppercase bases are conserved in all but one of the color-coded test species: *Monomorium pharaonis, Atta cephalotes, Vollenhovia emeryi, Acromyrmex echinatior, Lasius niger, Pogonomyrmex barbatus, Wasmannia auropunctata, Cardiocondyla obscurior or Linepithema humile*.