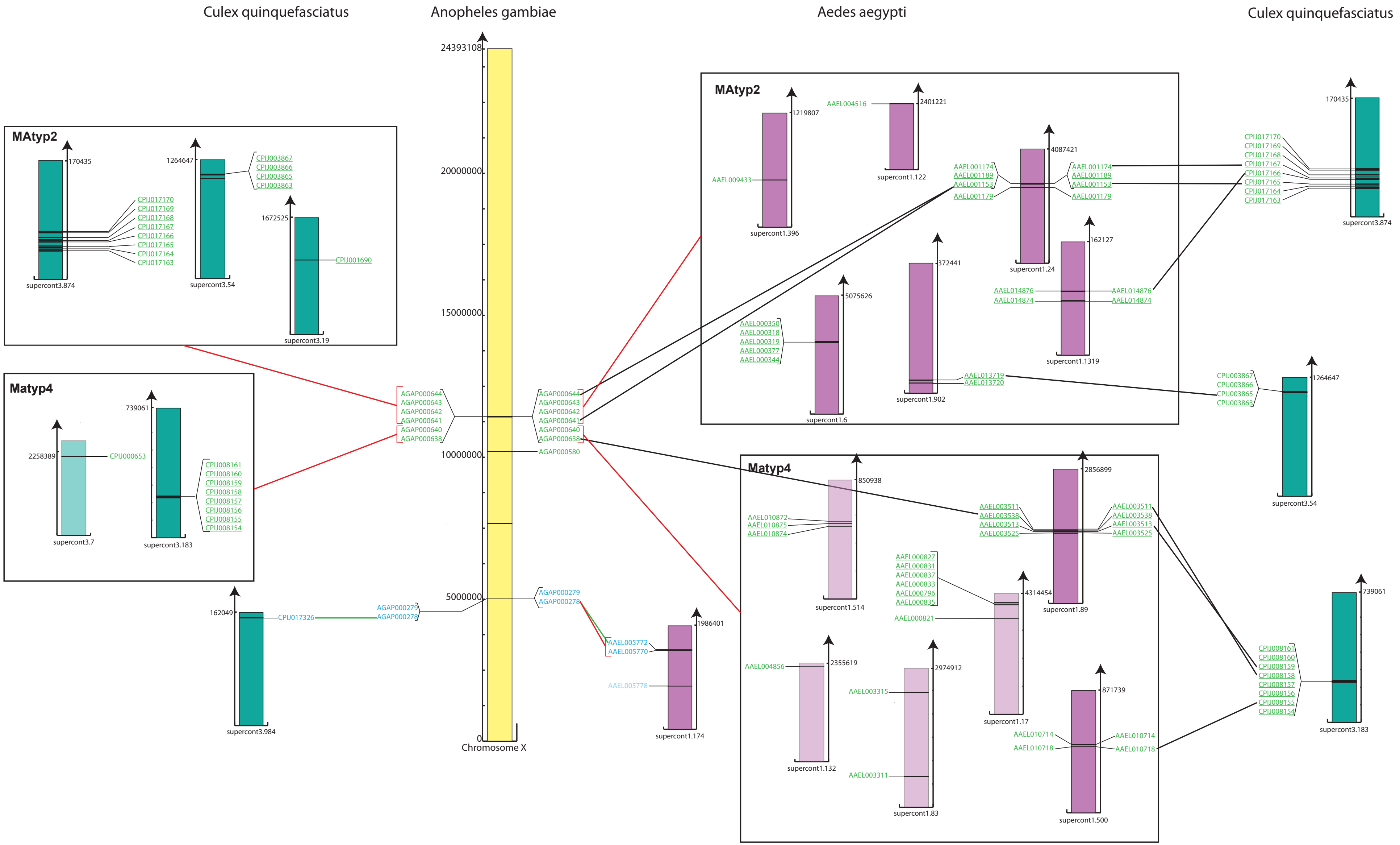
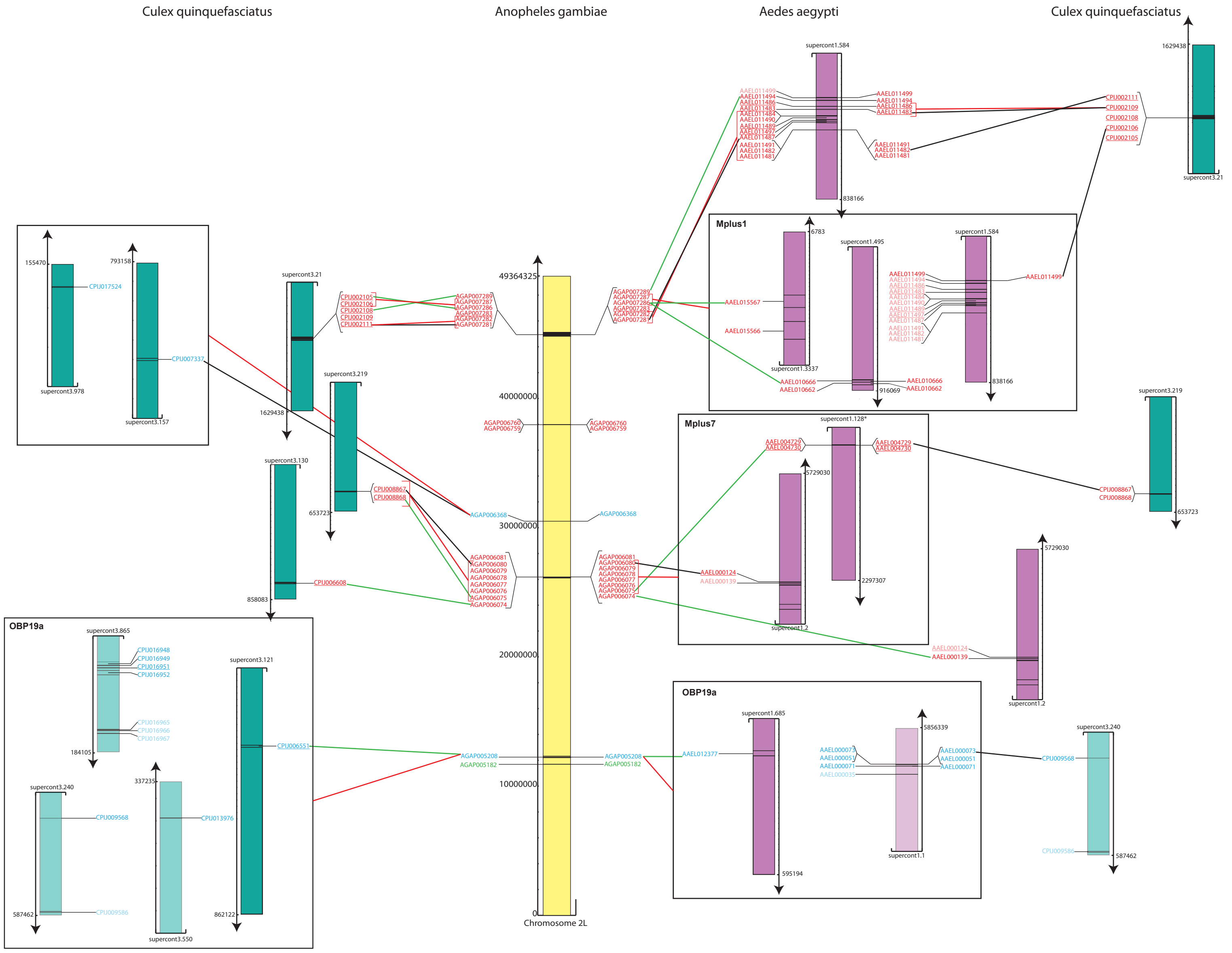
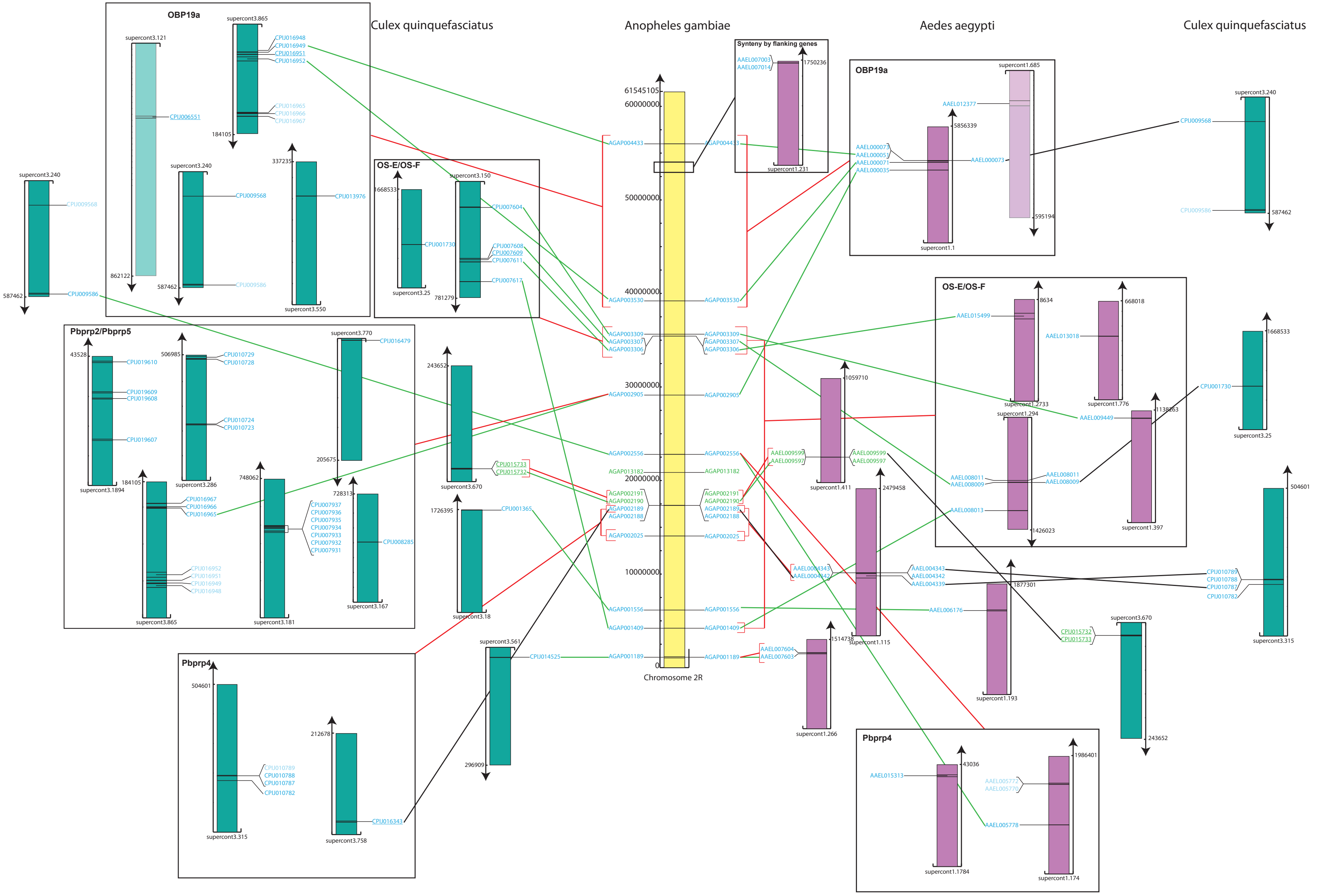


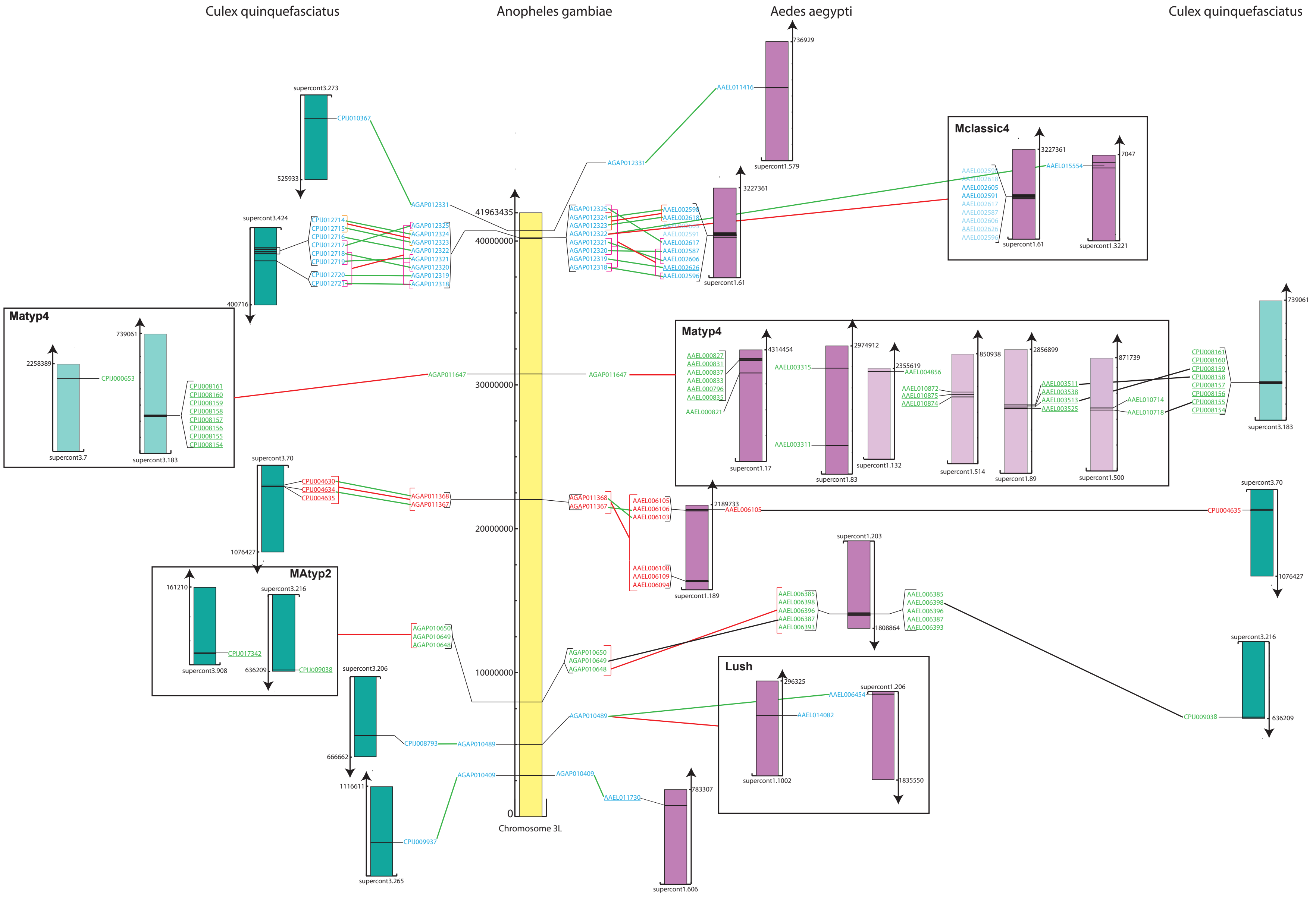
Supplementary figure 1a. Analysis of OBP genes distribution on X chromosome (yellow bar) of *Anopheles gambiae* and their corresponding orthologues & paralogues on *Aedes aegypti* (purple bars) and *Culex quinquefasciatus* (green) supercontigs. See additional legend for more details.



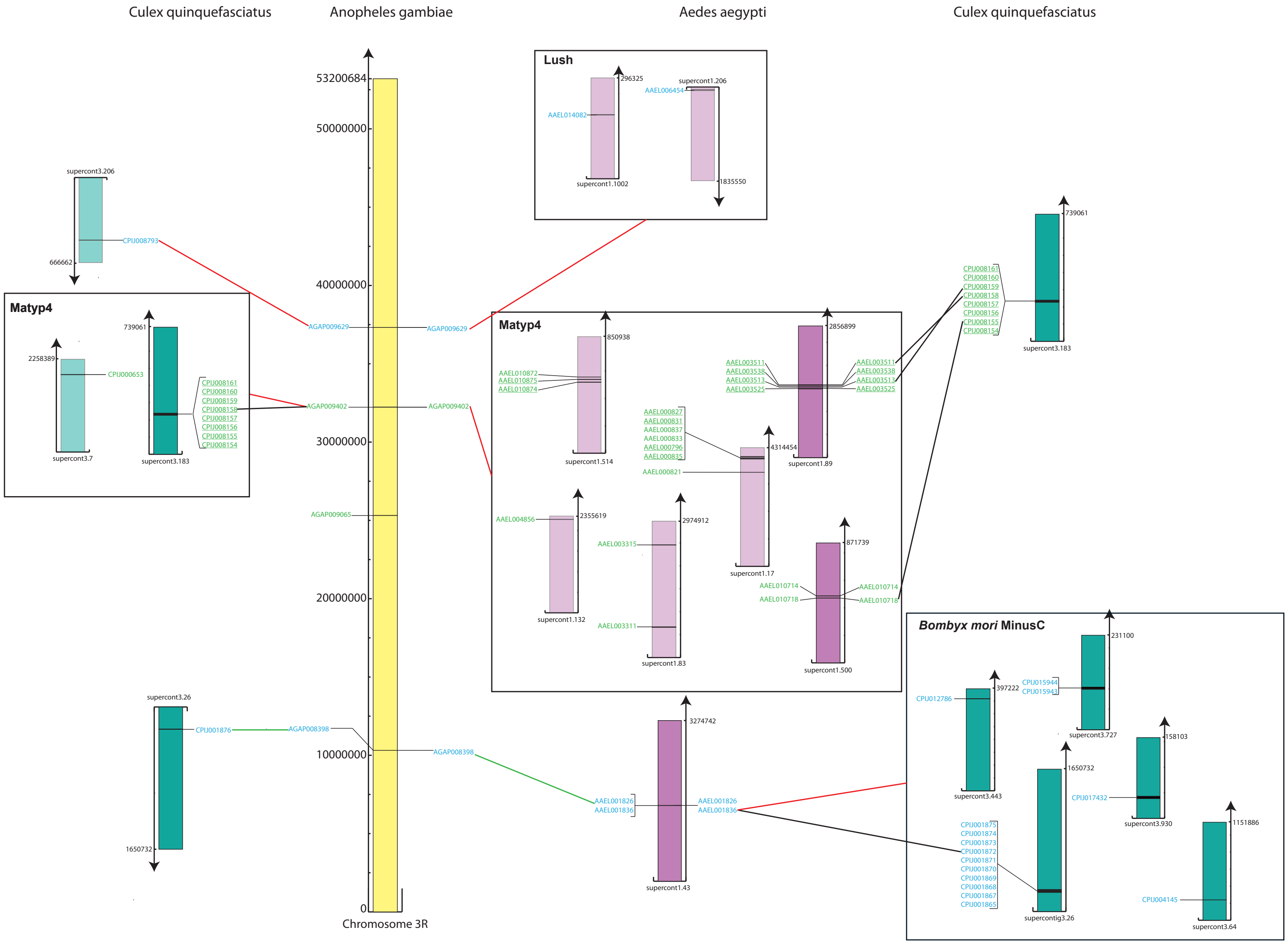
Supplementary figure 1b. Analysis of OBP genes distribution on 2L chromosome (yellow bar) of *Anopheles gambiae* and their corresponding orthologues & paralogues on *Aedes aegypti* (purple bars) and *Culex quinquefasciatus* (green bars) supercontigs. See additional legend for more details.



Supplementary figure 1c. Analysis of OBP genes distribution on 2R chromosome (yellow bar) of *Anopheles gambiae* and their corresponding orthologues & paralogues on *Aedes aegypti* (purple bars) and *Culex quinquefasciatus* (green bars) supercontigs. See additional legend for more details.

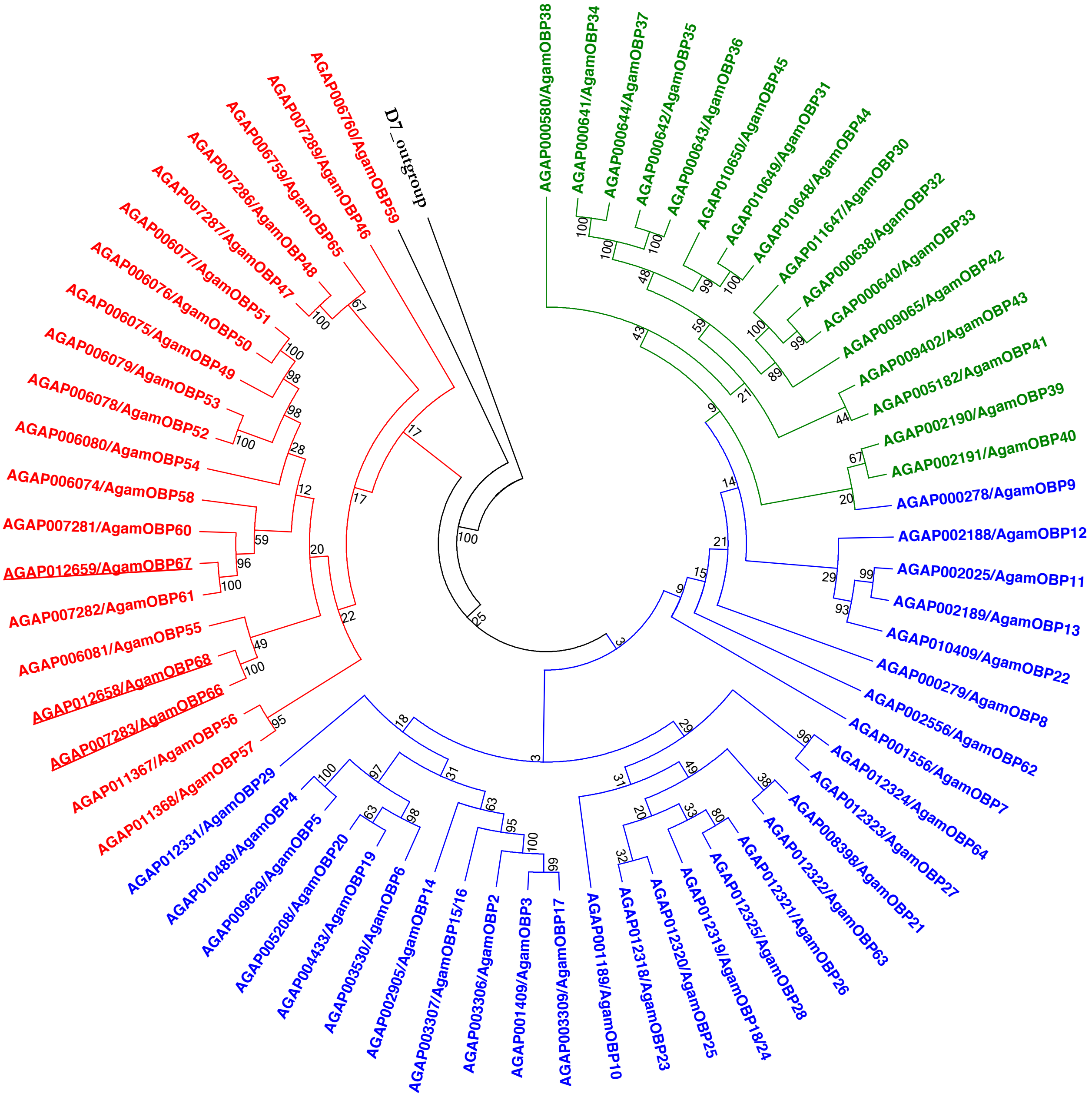


Supplementary figure 1d. Analysis of OBP genes distribution on 3L chromosome (yellow bar) of *Anopheles gambiae* and their corresponding orthologues & paralogues on *Aedes aegypti* (purple bars) and *Culex quinquefasciatus* (green bars) supercontigs. See additional legend for more details.

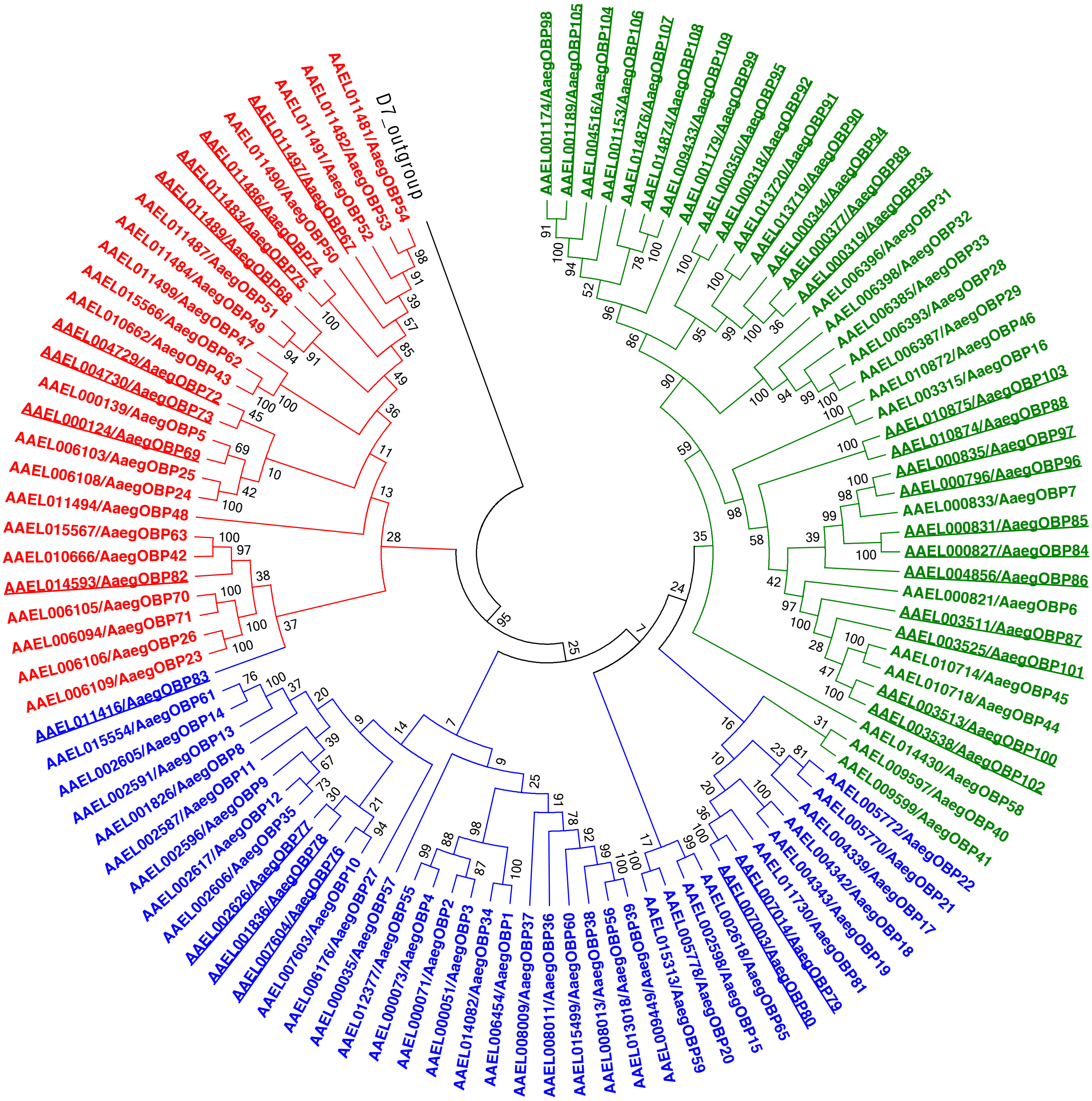


Supplementary figure 1e. Analysis of OBP genes distribution on 3R chromosome (yellow bar) of *Anopheles gambiae* and their corresponding orthologues & paralogues on *Aedes aegypti* (purple bars) and *Culex quinquefasciatus* (green bars) supercontigs. See additional legend for more details.

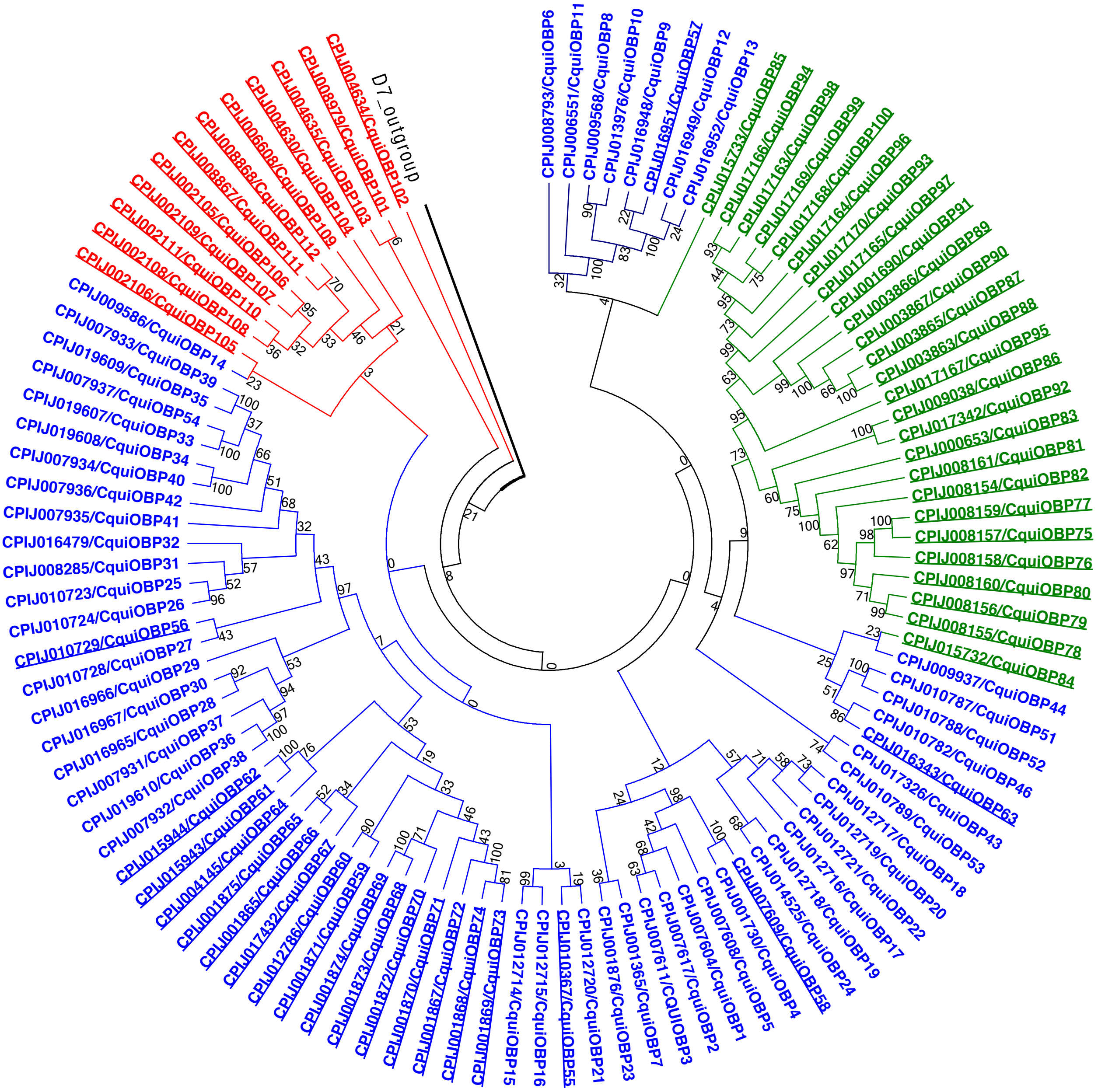
Additional legend to supplementary Figures 1a-e. Analysis of OBP genes distributions in *Anopheles gambiae*, *Aedes aegypti* and *Culex quinquefasciatus* genomes. Genes from the subfamilies of the OBP group are colored differently: *Classic* OBPs genes are printed in blue, *Atypical* OBPs in green and *PlusC* in red. The *An. gambiae* chromosomes are in yellow and are centrally located in the diagrams. The *Ae. aegypti* and *C. quinquefasciatus* super contigs are featured in purple and green respectively. Orthology between OBP genes was mainly established using the reverse blast hit (*rbh*) methodology (see materials and method for details). Paralogous relationships were confirmed through examination of the corresponding entries in the *inParanoid* database. Two-way orthologous relationships *i.e* only between genes in two genomes are connected with black lines while three-way orthologous relationships are featured using green lines. Red lines indicate inparalogous links between the connected sets of genes. The contigs from *C. quinquefasciatus* or *A. aegypti* are grouped in a square when all the enclosed OBP genes are from the same phylogenetic subcluster and are inparalogues, except for genes or contigs that are colored in semi-transparency : in these cases, the genes or contigs are displayed to recall their existence in the given cluster but do not share inparalogy relationship with the other enclosed OBP genes ; they might be orthologous to other genes in other chromosomal location. Underlined genes are newly identified genes in this work.



Supplementary Figure 2a. Rooted phylogenetic tree of the odorant binding proteins in the *Anopheles gambiae* genome. The *Classic* OBPs subfamily are colored blue, *Atypical* OBPs are colored green and *PlusC* OBPs are colored red. The bootstrap values of the branches are indicated on the nodes in percentage values. The names of identified clusters are indicated on the branches (see Figures 4-6).

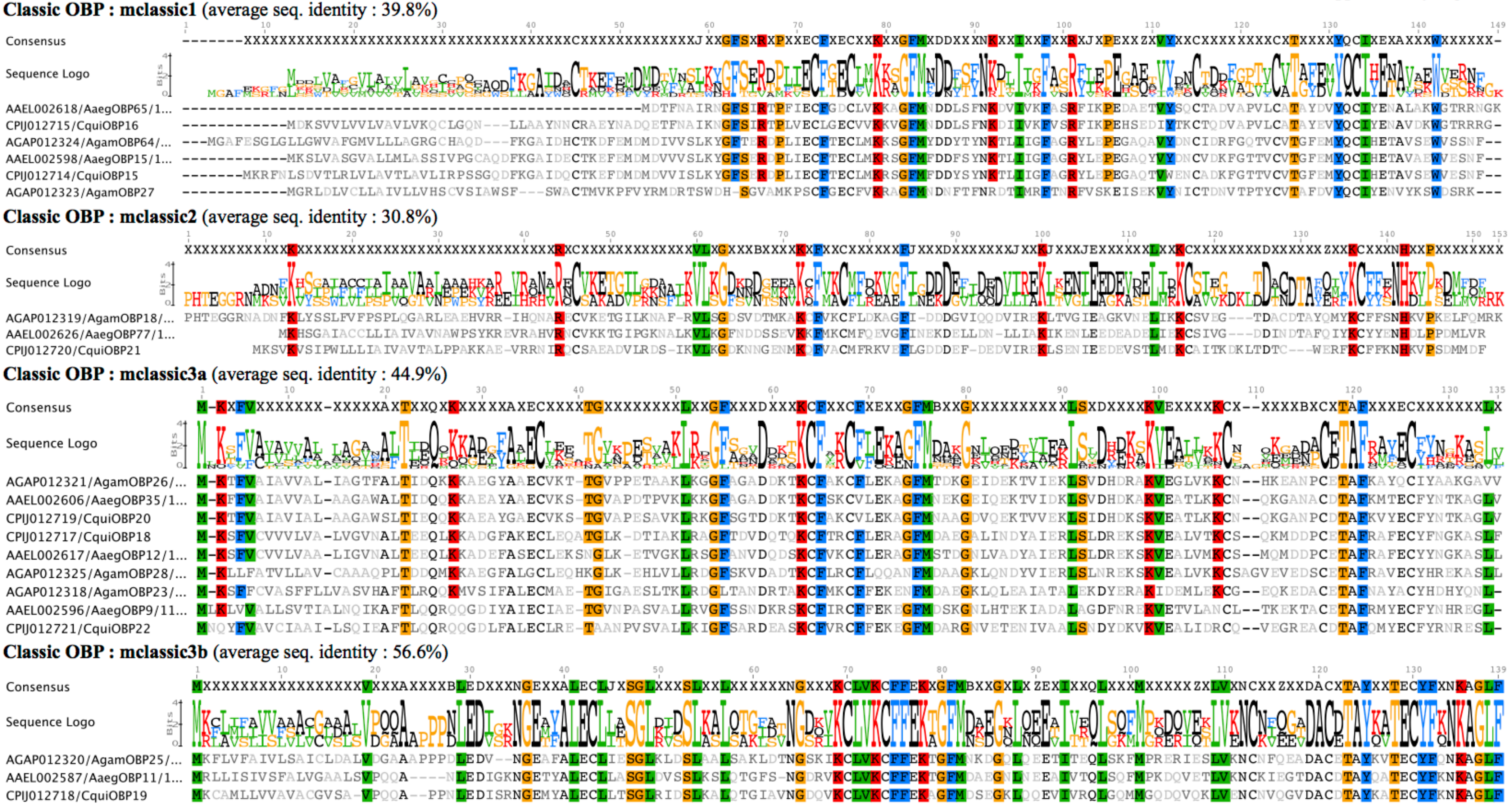


Supplementary Figure 2b. Rooted phylogenetic tree of the odorant binding proteins in the *Aedes aegypti* genome. The *Classic* OBPs subfamily are colored blue, *Atypical* OBPs are colored green and *PlusC* OBPs are colored red. The bootstrap values of the branches are indicated on the nodes in percentage values. The names of identified clusters are indicated on the branches (see Figures 4-6).



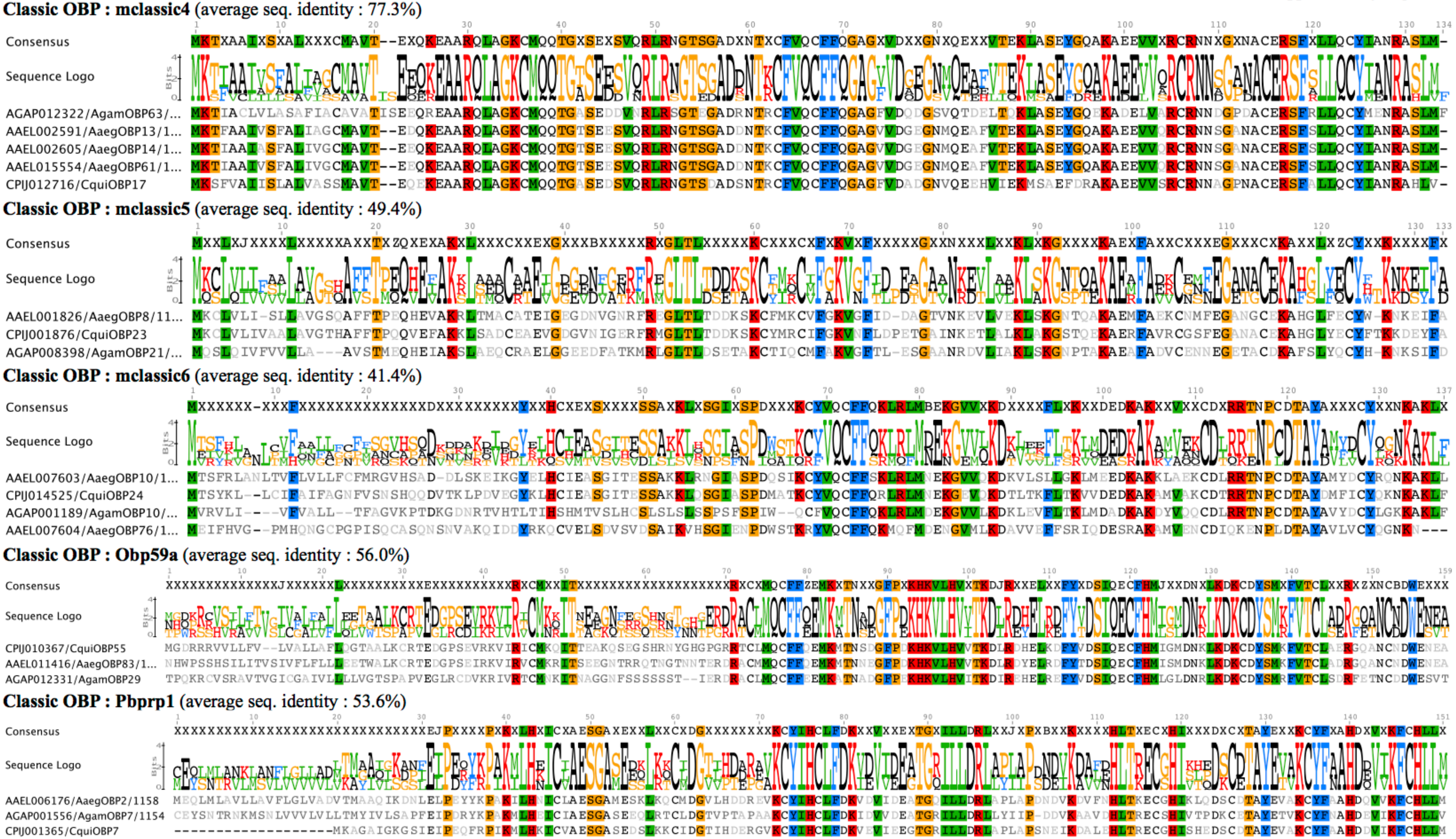
Supplementary Figure 2c. Rooted phylogenetic tree of the odorant binding proteins in the *Culex quinquefasciatus* genome. The *Classic* OBPs subfamily are colored blue, *Atypical* OBPs are colored green and *PlusC* OBPs are colored red. The bootstrap values of the branches are indicated on the nodes in percentage values. The names of identified clusters are indicated on the branches (see Figures 4-6).

Supplementary Figure 3a

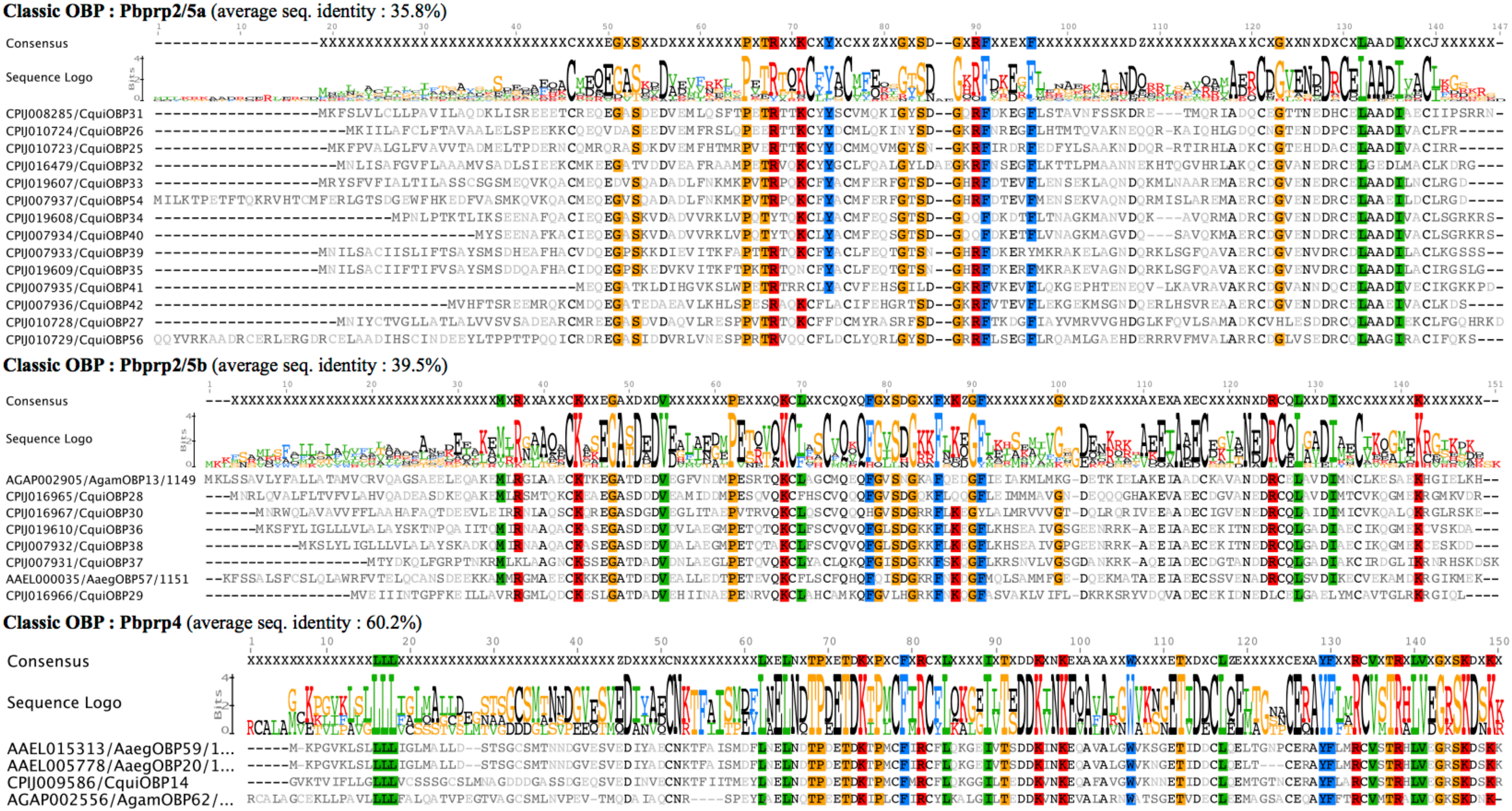


Supplementary Figure 3a. Protein sequence alignments of OBPs that belongs to the different clusters of the *Classic* subfamily. See additional legend for details.

Supplementary Figure 3a



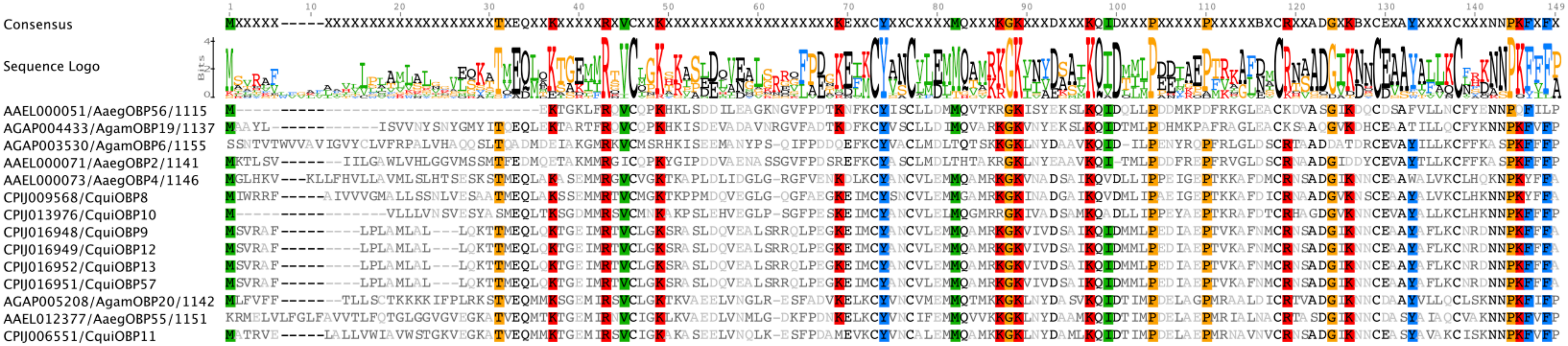
Supplementary Figure 3a. Protein sequence alignments of OBPs that belongs to the different clusters of the *Classic* subfamily. See additional legend for details.



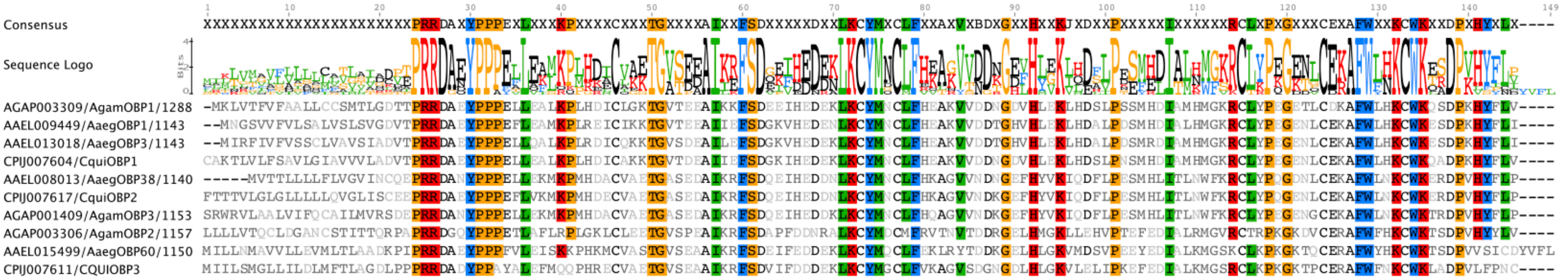
Supplementary Figure 3a. Protein sequence alignments of OBPs that belongs to the different clusters of the *Classic* subfamily. See additional legend for details.

Classic OBP : LUSH (average seq. identity : 44.7%)

Classic OBP : Obp19a (average seq. identity : 45.9%)



Classic OBP : OS-E/OS-F(a) (average seq. identity : 55.2%)



Supplementary Figure 3a. Protein sequence alignments of OBPs that belongs to the different clusters of the *Classic* subfamily. See additional legend for details.

Supplementary Figure 3a


Classic OBP : OS-E/OS-F(b) (average seq. identity : 36.7%)

Classic OBP : Bombyx mori (minus C) (average seq. identity : 21.5%)

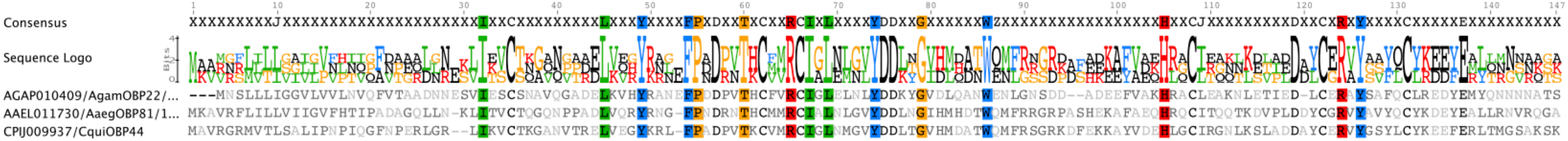
Classic OBP : Obp99a (minus C) (average seq. identity : 72.9%)

Classic OBP : mclassic7 (average seq. identity : 46.0%)

Supplementary Figure 3a. Protein sequence alignments of OBPs that belongs to the different clusters of the *Classic* subfamily. See additional legend for details.

Supplementary Figure 3a

Classic OBP : mclassic8 (average seq. identity : 30.5%)



Classic OBP : mclassic9a (average seq. identity : 40.6%)

Classic OBP : mclassic9b (average seq. identity : 25.1%)

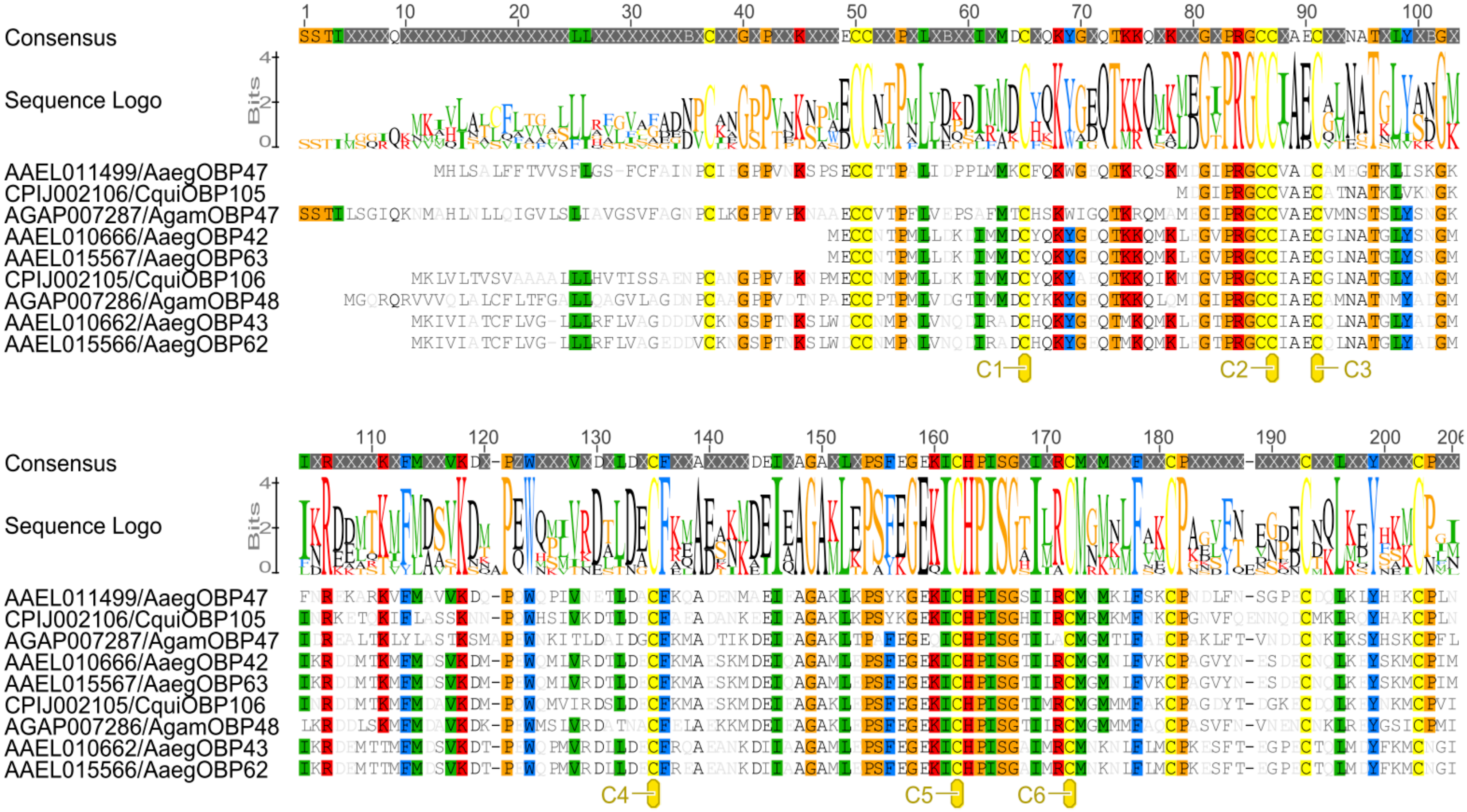
Additional legend to Figure 3a.

Shown are the alignments for the *Classic* OBPs from the three mosquito genomes *Anopheles gambiae, Aedes aegypti* and *Culex quinquefasciatus*. These were obtained after aligning the sequences with a structure-based profile using ClustalX software. The global *Classic* OBP alignment was splitted into different parts corresponding to the phylogenetic clusters that is established in this study (see Figures 4 and 5). The residues that are highly conserved (with 75% or more degree of conservation) are highlighted in the sequences and in the sequence logos above the alignment. The consensus sequence is also featured on top of each alignment. The average pairwise sequence identities within each cluster are indicated. These diagrams were generated by the Geneious software.

Supplementary Figure 3a. Protein sequence alignments of OBPs that belongs to the different clusters of the *Classic* subfamily. See additional legend for details.

Supplementary Figure 3b.

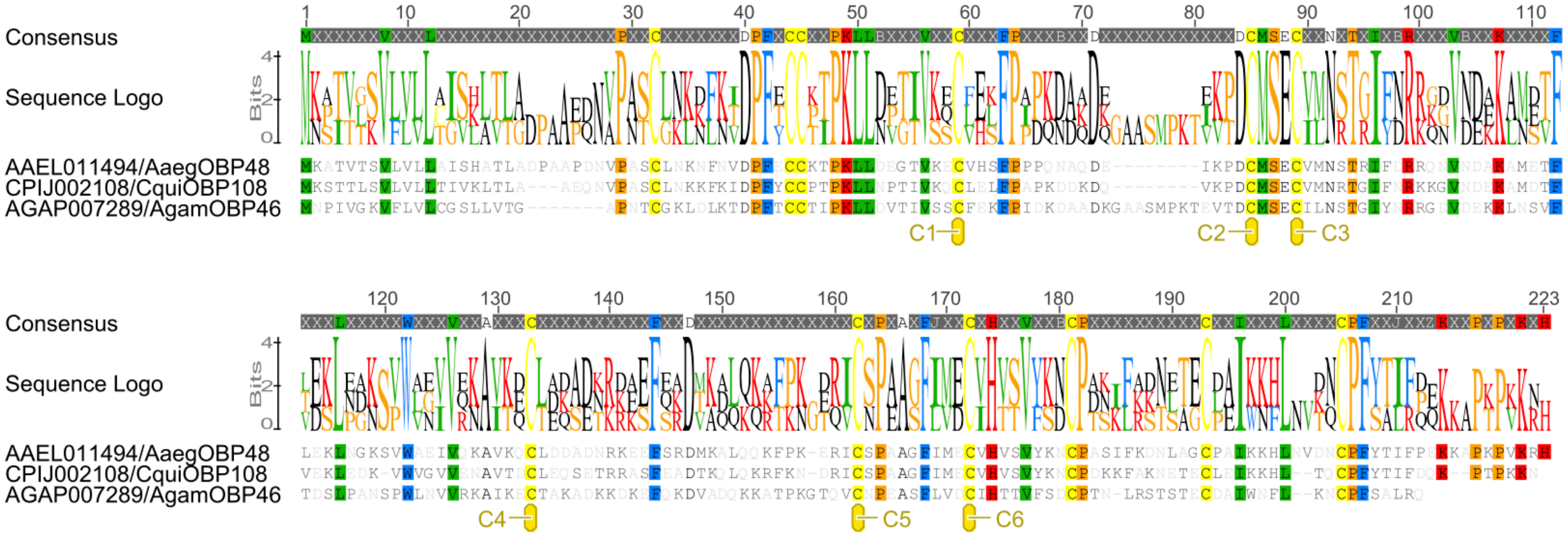
Plus C : mplus1 (average seq. identity : 58.6%)



Supplementary Figure 3b. Protein sequence alignments of OBPs that belongs to the different clusters of the *PlusC* subfamily. See additional legend for details.

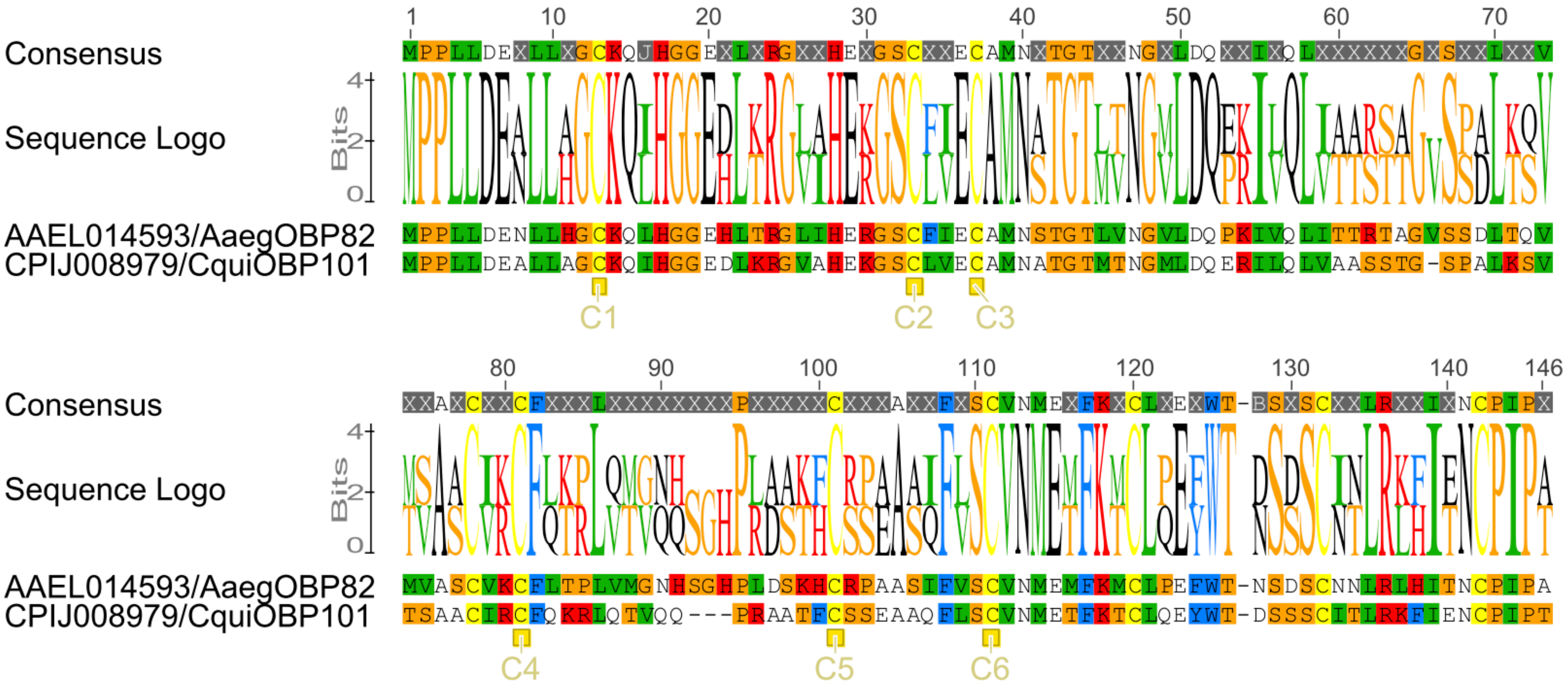
Supplementary Figure 3b.

Plus C : mplus2 (average seq. identity : 41.3%)



Supplementary Figure 3b. Protein sequence alignments of OBPs that belongs to the different clusters of the *PlusC* subfamily. See additional legend for details.

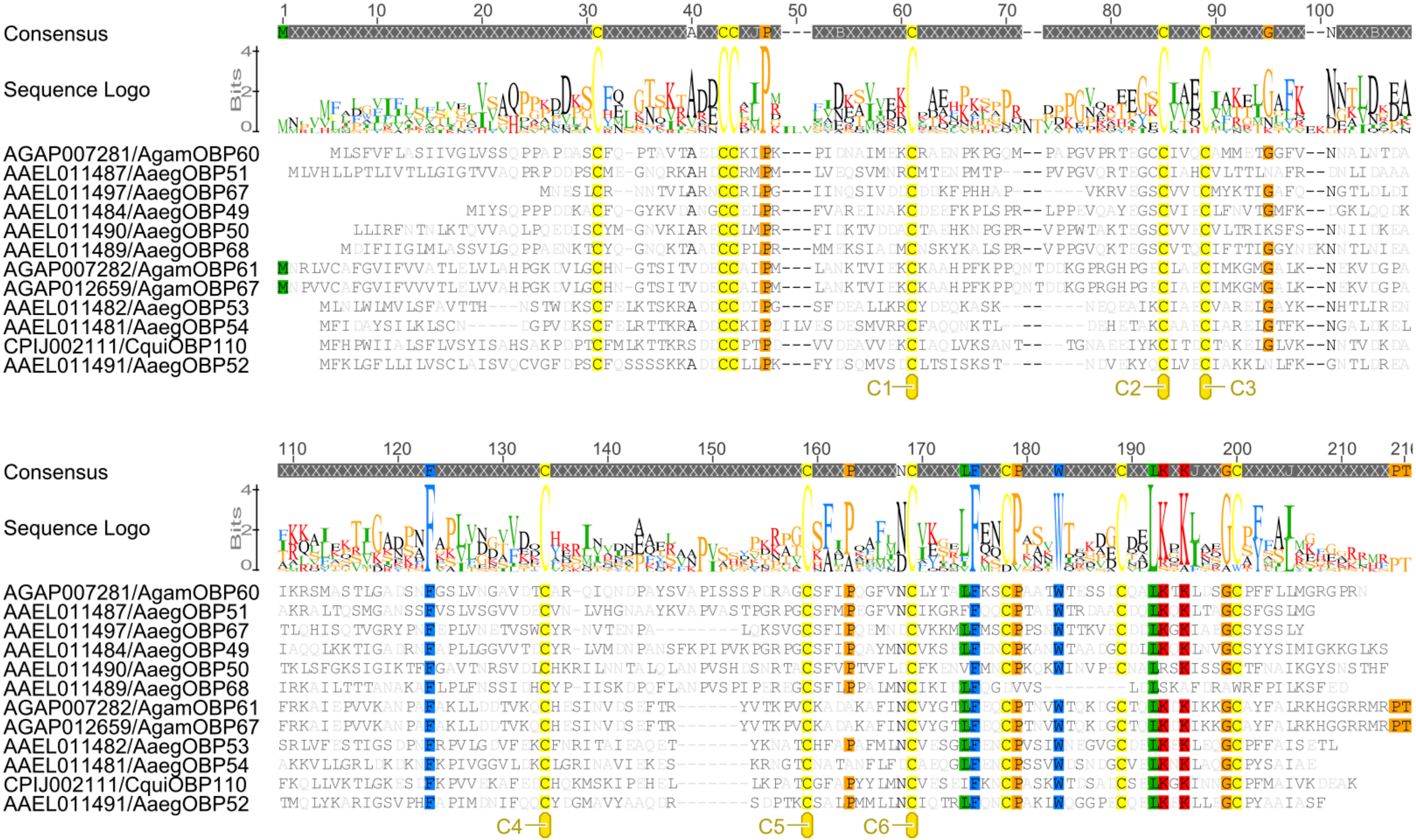
Plus C : mplus3 (average seq. identity : 53.8%)



Supplementary Figure 3b. Protein sequence alignments of OBPs that belongs to the different clusters of the *PlusC* subfamily. See additional legend for details.

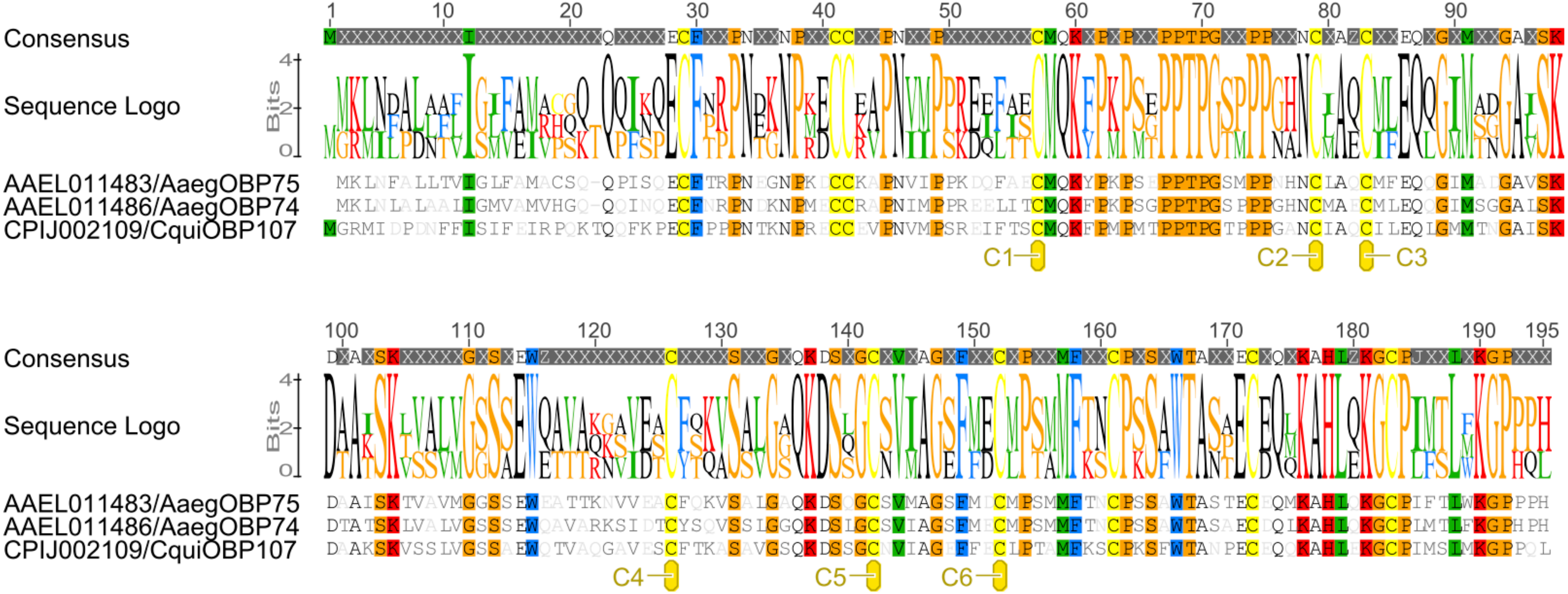
Supplementary Figure 3b.

Plus C : mplus4 (average seq. identity : 28.0%)



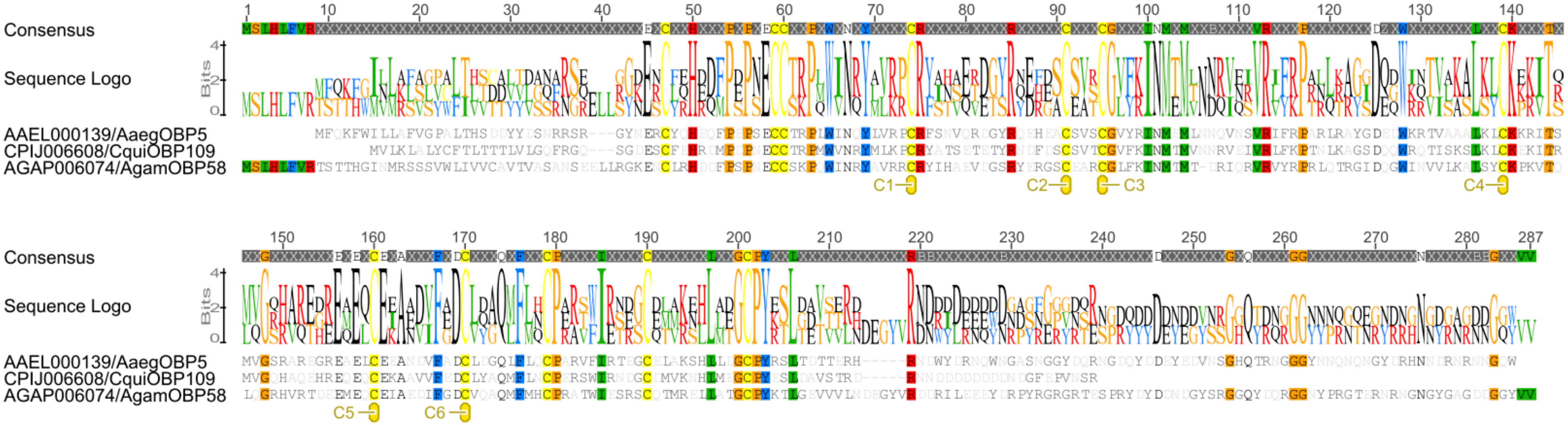
Supplementary Figure 3b. Protein sequence alignments of OBPs that belongs to the different clusters of the *PlusC* subfamily. See additional legend for details.

Plus C : mplus5 (average seq. identity : 56.8%)



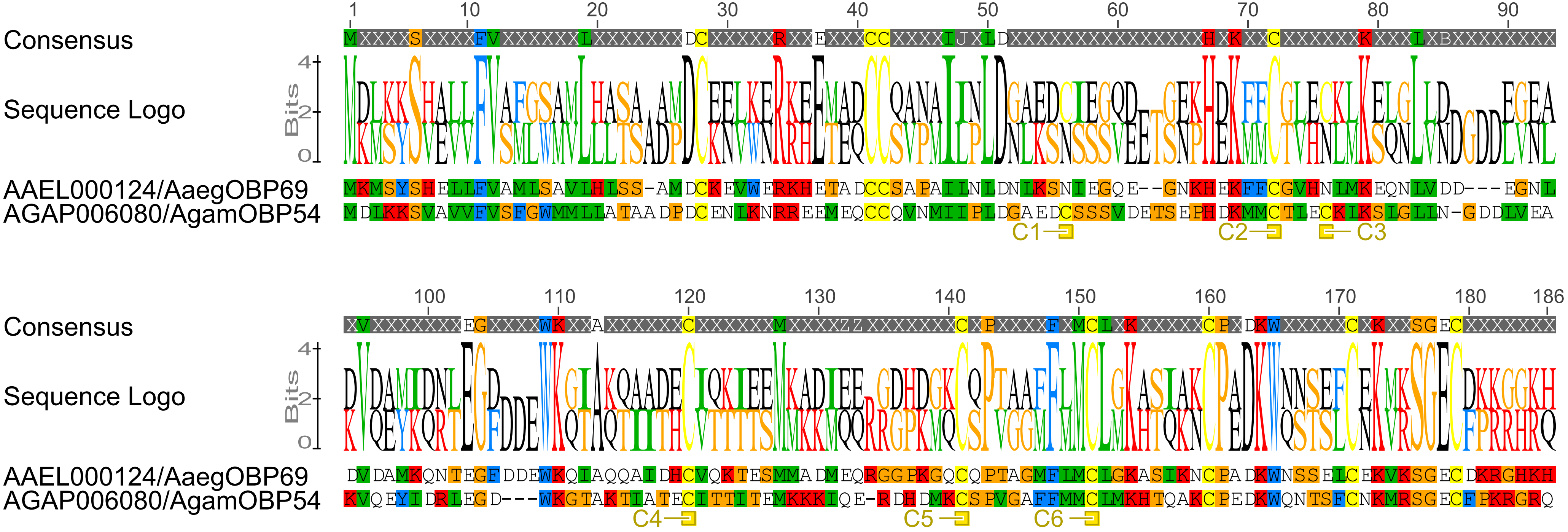
Supplementary Figure 3b. Protein sequence alignments of OBPs that belongs to the different clusters of the *PlusC* subfamily. See additional legend for details.

Plus C : mplus6 (average seq. identity : 33.8%)



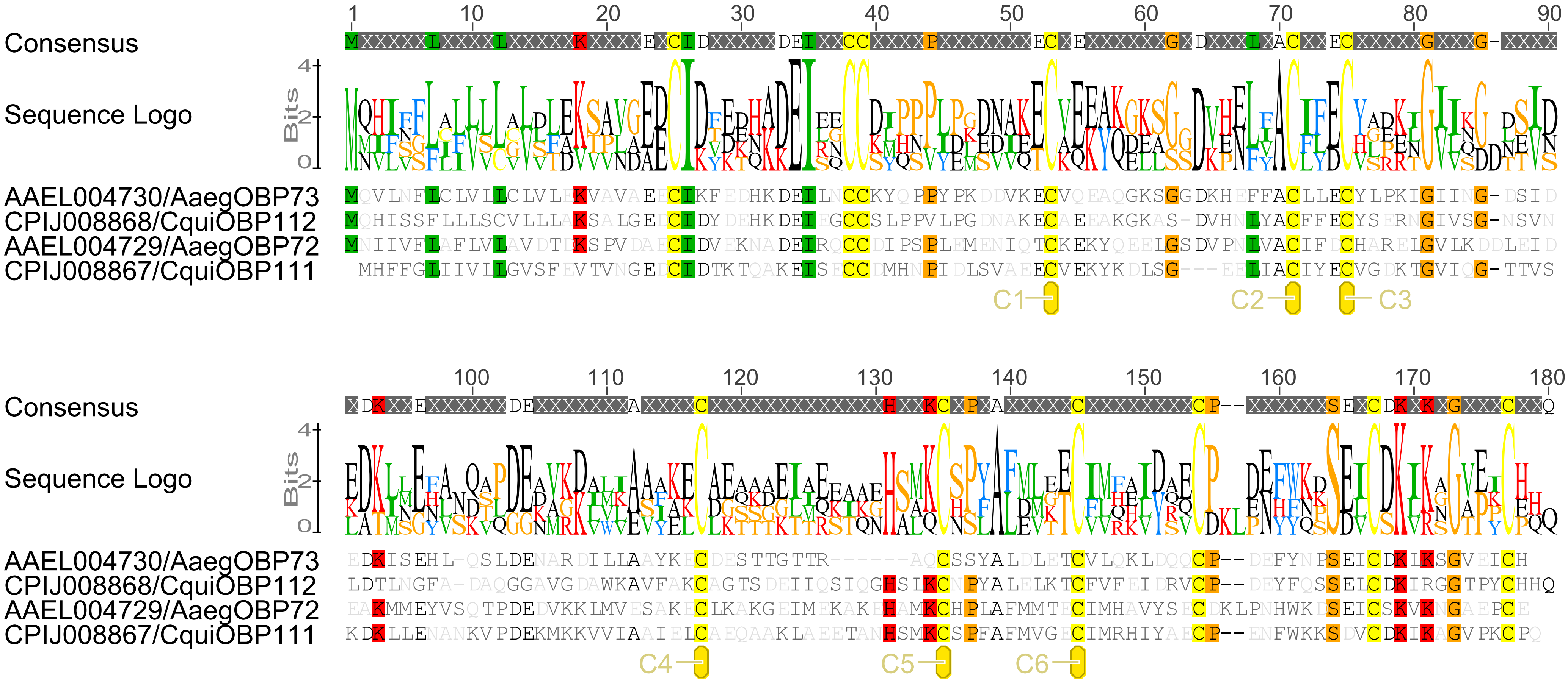
Supplementary Figure 3b. Protein sequence alignments of OBPs that belongs to the different clusters of the *PlusC* subfamily. See additional legend for details.

Plus C : mplus7 (average seq. identity : 24.2%)



Supplementary Figure 3b. Protein sequence alignments of OBPs that belongs to the different clusters of the *PlusC* subfamily. See additional legend for details.

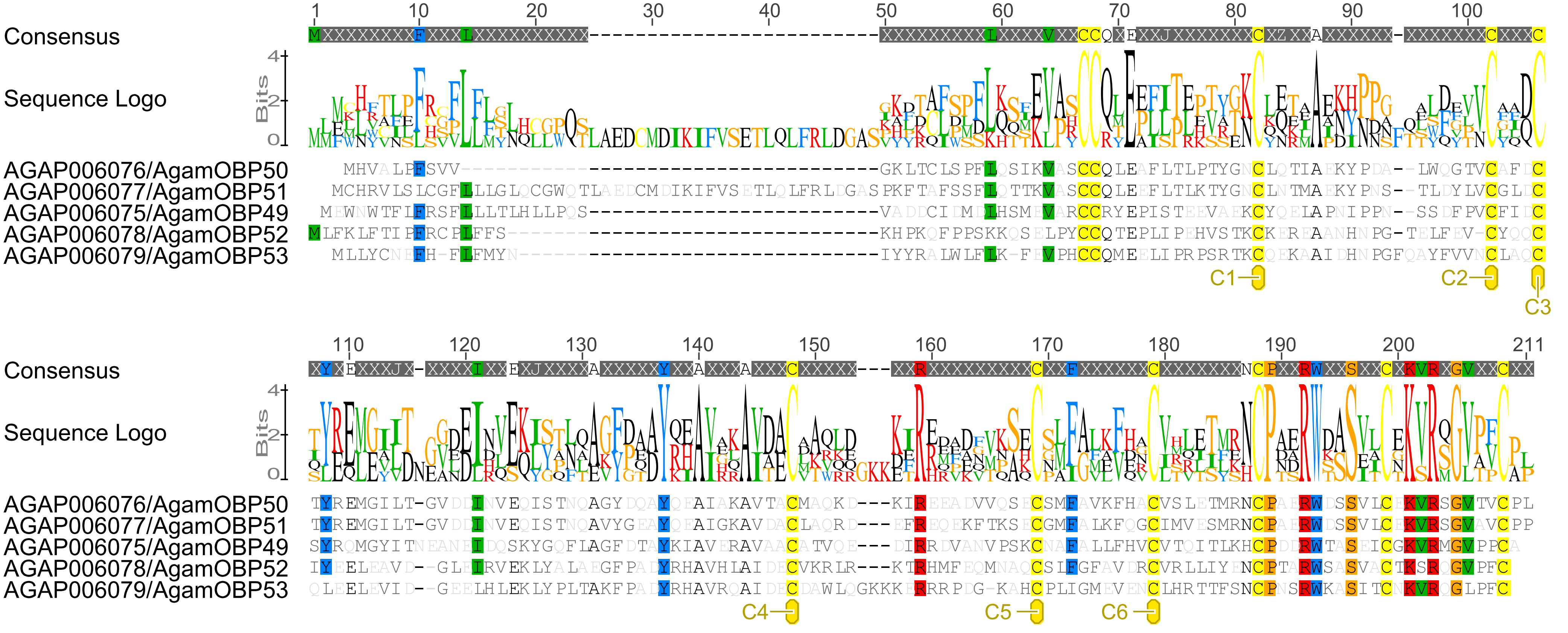
Plus C : mplus8 (average seq. identity : 29.5%)



Supplementary Figure 3b. Protein sequence alignments of OBPs that belongs to the different clusters of the *PlusC* subfamily. See additional legend for details.

Supplementary Figure 3b.

Plus C : mplus9 (average seq. identity : 29.7%)

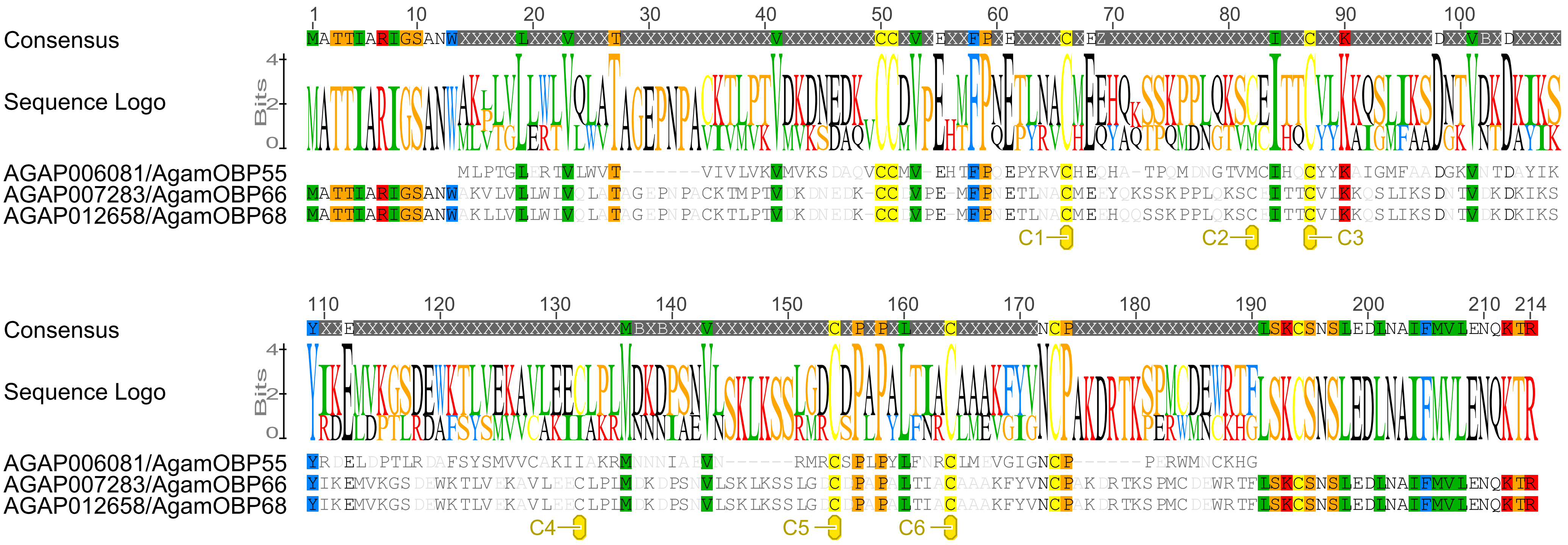


```
Consensus             MXXXXXXXXFXXXLXXXXXXXXXX-------------------------XXXXXXXXXLXXXXVXXCCQXEXXJXXXXXXXCXZXXAXXXXXX-XXXXXXXCXXXC
AGAP006076/AgamOBP50     MHVALPFSVV------------------------------------GKLTCLSPFLQSIKVASCCQLEAFLTLPTYGNCLQTIAEKYPDA--LWQGTVCAFDC
AGAP006077/AgamOBP51    MCHRVLSLCGFLLLGLQCGWQTLAEDCMDIKIFVSETLQLFRLDGASPKFTAFSSFLQTTKVASCCQLEEFLTLKTYGNCLNTMAEKYPNS--TLDYLVCGLDC
AGAP006075/AgamOBP49   MEWNWTFLFRSFLLLTLHLLPQS-------------------------VADDCIDMDLHSMEVARCCRYEPISTEEVAEKCYQELAPNIPPN--SSDFPVCFIDC
AGAP006078/AgamOBP52  MLFKLFTIPFRCPLFFS--------------------------------KHPKQFPPSKKQSELPYCCQTEPLIPEHVSTKCKEREAANHNPG-TELFEV-CYQQC
AGAP006079/AgamOBP53    MLLYCNEFH-FLFMYN-------------------------------IYYRALWLFLK-FEVPHCCQMEELIPRPSRTKCQEKAAIDHNPGFQAYFVVNCLAQC

Consensus             XYXEXXXJX-XXXXIXXEXJXXXXAXXXXXYXXAXXXAXXXCXXXXX---XXRXXXXXXXXXXCXXFXXXXXXXCXXXXXXXNCPXXRWXXSXXCXKVRXGVXXCXX
AGAP006076/AgamOBP50  TYREMGILT-GVDDINVEQISTNQAGYDQAYQEAIAKAVTACMAQKD---KIREEADVVQSECSMFAVKFHACVSLETMRNCPAERWDSSVLCEKVRSGVTVCPL
AGAP006077/AgamOBP51  TYREMGILT-GVDDINVEQISTNQAVYGEAYQEAIGKAVDACLAQRD---EFREQEKFTKSECGMFALKFQGCIMVESMRNCPAERWDSSVLCEKVRSGVAVCPP
AGAP006075/AgamOBP49  SYRQMGYITNEANEIDQSKYGQFLAGFDTAYKIAVERAVAACATVQE---DIRRDVANVPSKCNAFALLFHVCVTQITLKHCPDDRWTASEICGKVRMGVPPCA
AGAP006078/AgamOBP52  IYEELEAVD--GLEIRVEKLYALAEGFPADYRHAVHLAIDECVKRLR---KTRHMFEQMNAQCSLFGFAVDRCVRLLIYENCPTARWSASVACTKSRQGVPFC
AGAP006079/AgamOBP53  QLEELEVID--GEELHLEKLYPLTAKFPADYRHAVRQAIDECDAWLQGKKKERRRPDG-KAHCPLIGMEVENCLHRTTFSNCPNSRWKASITCNKVRQGLPFC
```

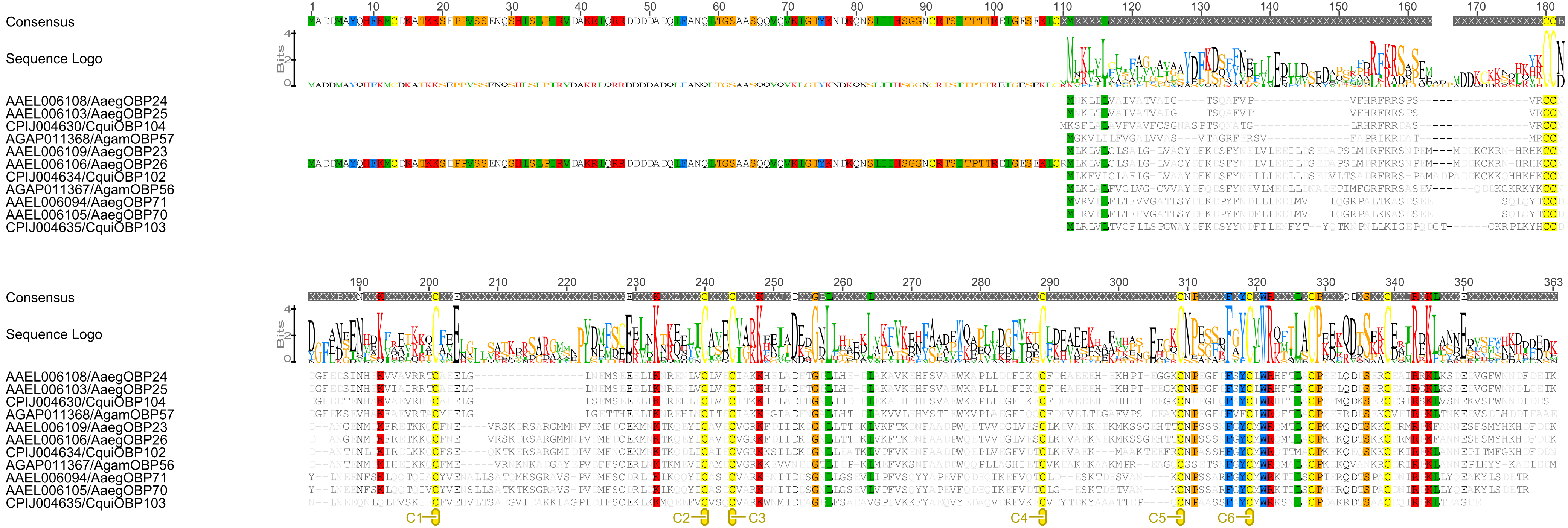
Supplementary Figure 3b. Protein sequence alignments of OBPs that belongs to the different clusters of the *PlusC* subfamily. See additional legend for details.

Plus C : mplus10 (average seq. identity : 48.1%)



Supplementary Figure 3b. Protein sequence alignments of OBPs that belongs to the different clusters of the *PlusC* subfamily. See additional legend for details.

Plus C : mplus11 (average seq. identity : 35.0%)



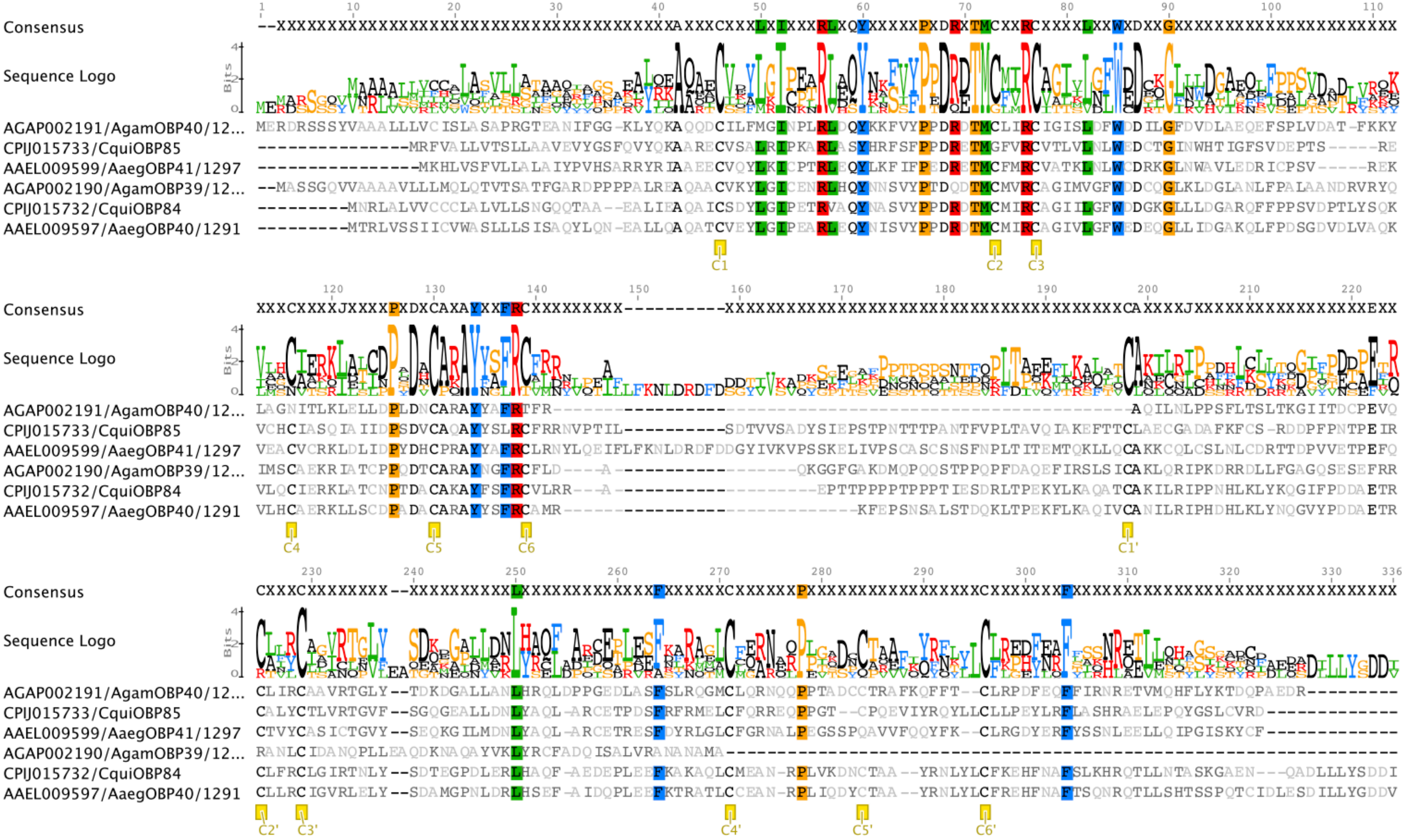
Additional legend to supplementary Figure 3b.

Shown are the alignments for the *PlusC* OBPs from the three mosquito genomes *Anopheles gambiae, Aedes aegypti* and *Culex quinquefasciatus*. These were obtained after aligning the sequences with a structure-based profile using ClustalX software. The global *PlusC* OBP alignment was splitted into different parts corresponding to the phylogenetic clusters that is established in this study (see Figures 4 and 5). The residues that are highly conserved (with 75% or more degree of conservation) are highlighted in the sequences and in the sequence logos above the alignment. The consensus sequence is also featured on top of each alignment. The average pairwise sequence identities within each cluster are indicated. These diagrams were generated by the Geneious software.

Supplementary Figure 3b. Protein sequence alignments of OBPs that belongs to the different clusters of the *PlusC* subfamily. See additional legend for details.

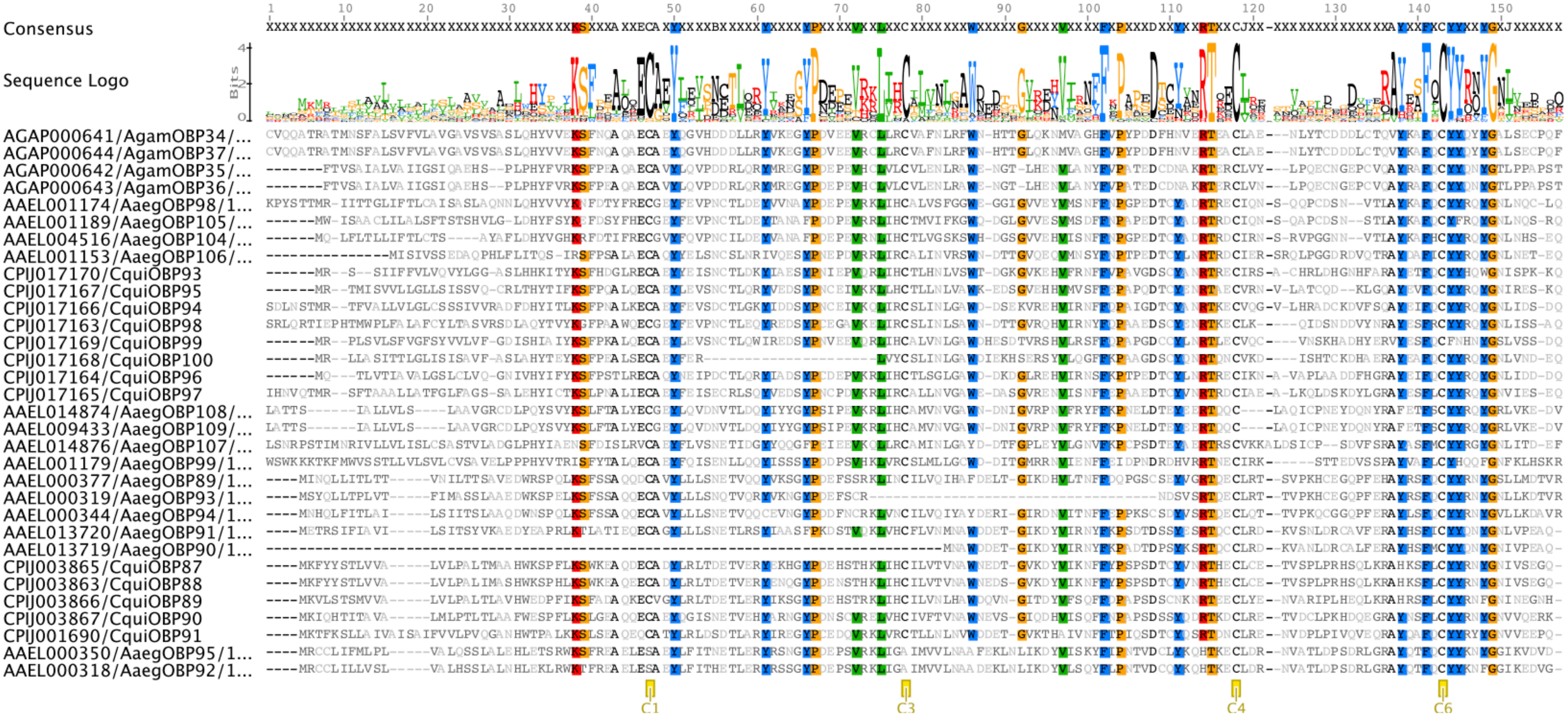
Supplementary Figure 3c

Atypical : matype1 (average seq. identity : 27.3%)



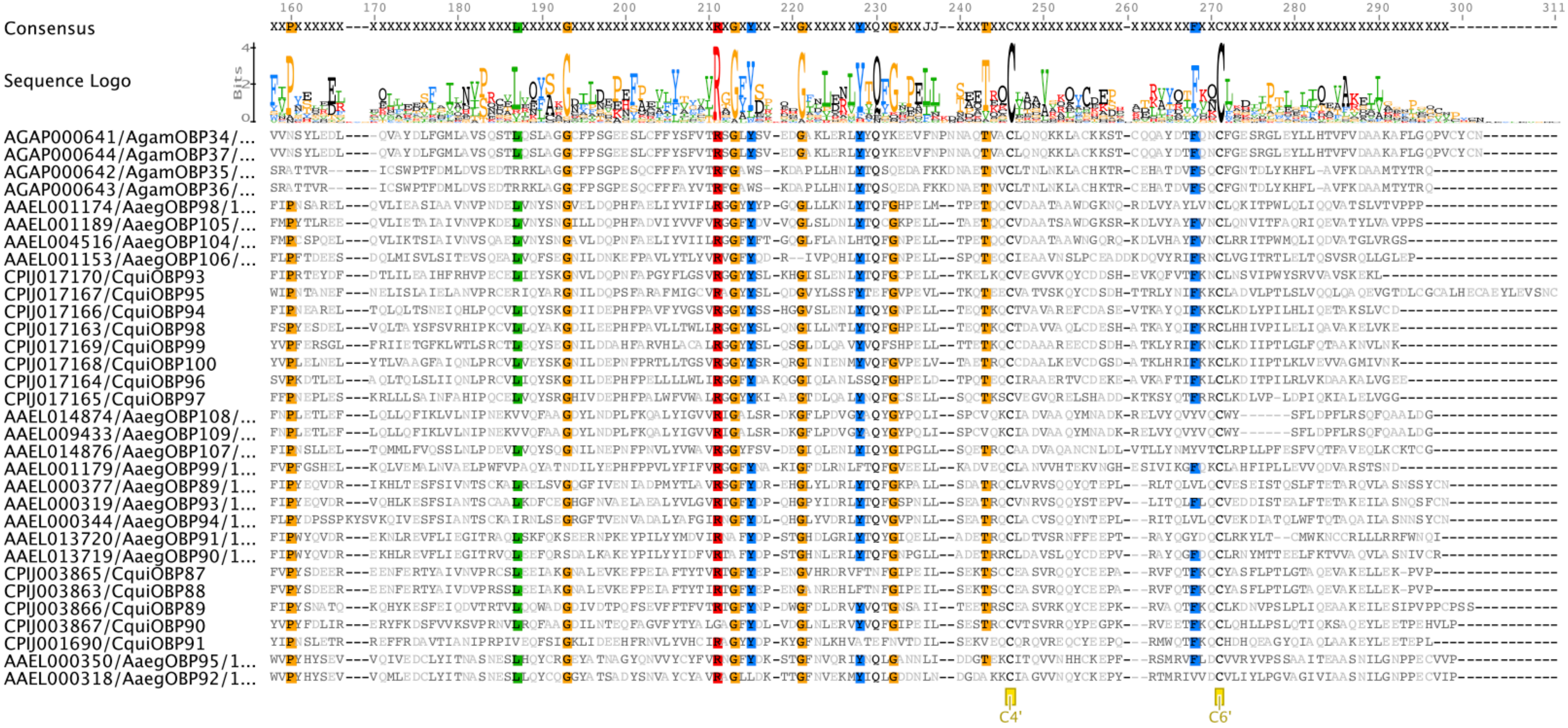
Supplementary Figure 2c. Protein sequence alignments of OBPs that belongs to the different clusters of the *Atypical* subfamily. See additional legend for details.

Atypical : matype2 (average seq. identity : 29.3%) - to be continued on next page



Supplementary Figure 3c. Protein sequence alignments of OBPs that belongs to the different clusters of the *Atypical* subfamily. See additional legend for details.

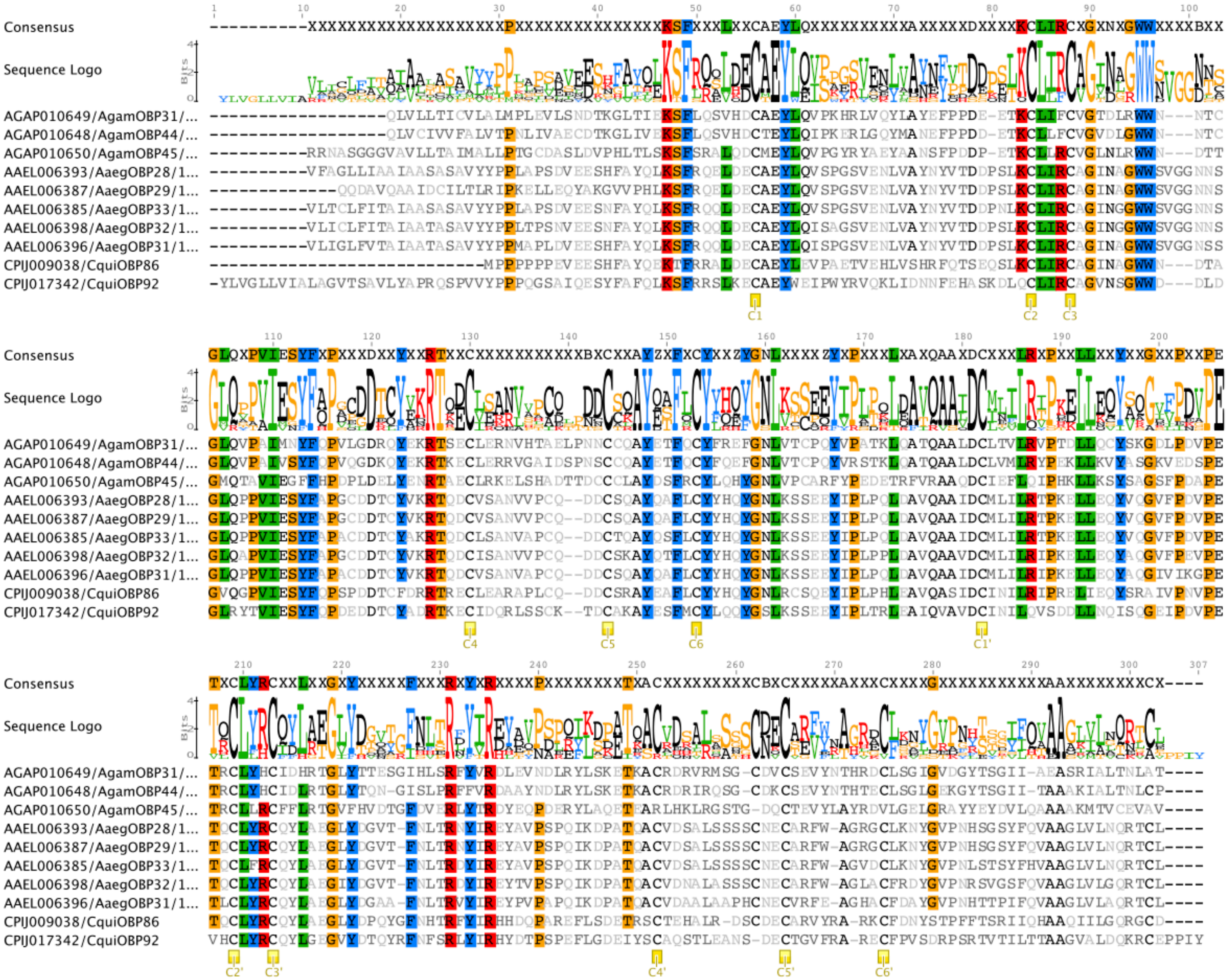
Atypical : matype2 (average seq. identity : 29.3%) - continued from previous page



Supplementary Figure 3c. Protein sequence alignments of OBPs that belongs to the different clusters of the *Atypical* subfamily. See additional legend for details.

Supplementary Figure 2c

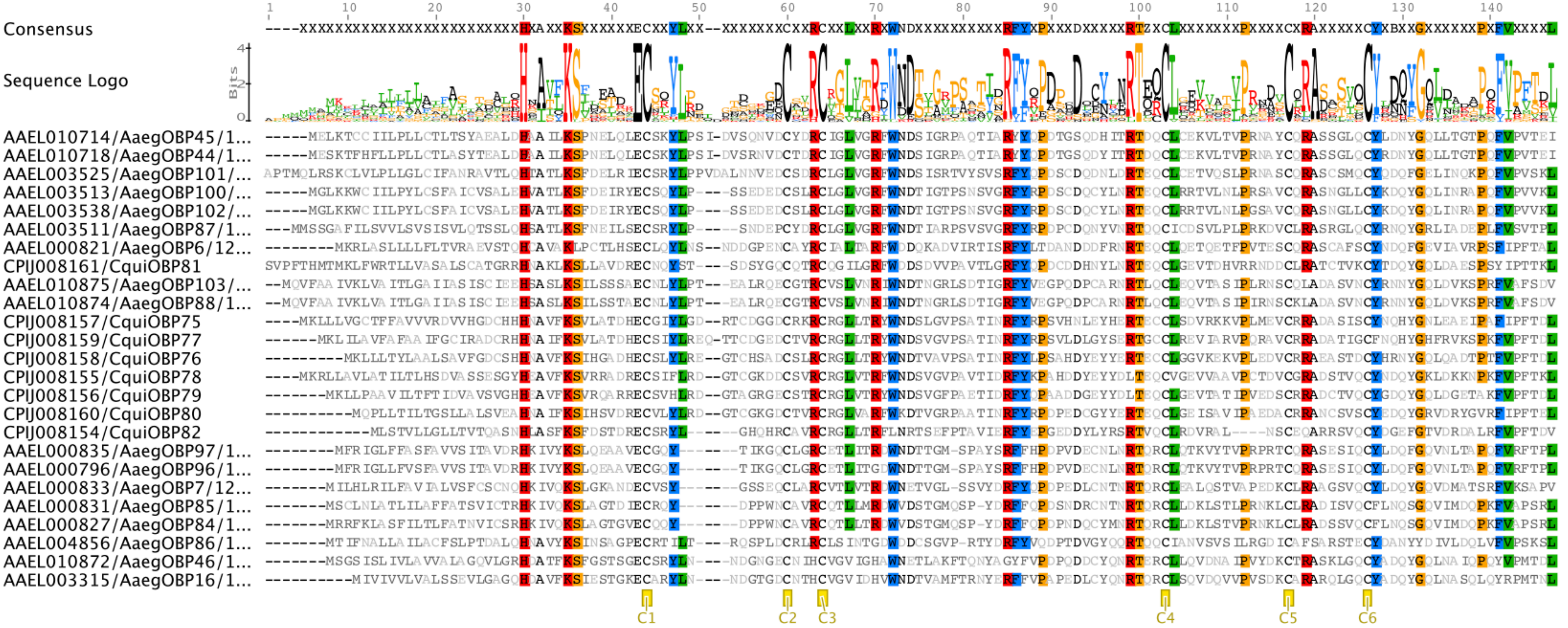
Atypical : matype3 (average seq. identity : 48.3%)



Supplementary Figure 3c. Protein sequence alignments of OBPs that belongs to the different clusters of the *Atypical* subfamily. See additional legend for details.

Supplementary Figure 2c

Atypical : matype4 (average seq. identity : 33.5%) - to be continued on next page

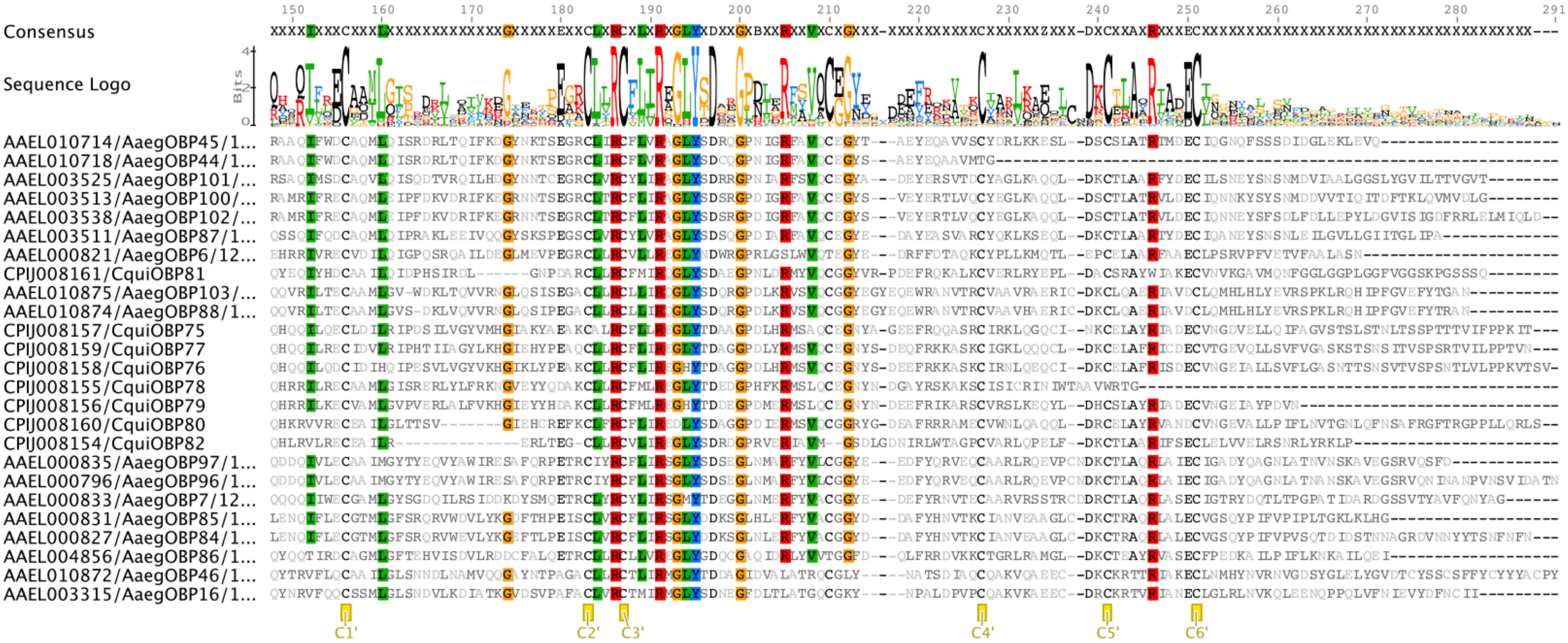


```
Consensus                    ----XXXXXXXXXXXXXXXXXXXXXXXXXHXAXXKSXXXXXXECXXYLXX--XXXXXXXCXXRCXXLXXRXWNDXXXXXXXXXXRFYXPXXXDXXXXXRTZXCLXXXXXXXPXXXXCXRAXXXXXCYXBXXGXXXXXXPXFVXXXXL
AAEL010714/AaegOBP45/1...    -----MELKTCCIILPLLCTLTSYAEALDHAAILKSPNELQLECSKYLPSI-DVSQNVDCYDRCIGLVGRFWNDSIGRPAQTIARYYQPDTGSQDHITRTDQCLCEKVLTVPRNAYCQRASSGLQCYLDNYGQLLTGTPQFVPVTEI
AAEL010718/AaegOBP44/1...    -----MESKTFHFLLPLLCTLASYTEALDHAAILKSPNELQLECSKYLPSI-DVSRNVDCTDRCIGLVGRFWNDSIGRPAQTIARYYQPDTGSQDYITRTDQCLCEKVLTVPRNAYCQRASSGLQCYRDNYGQLLTGTPQFVPVTEI
AAEL003525/AaegOBP101/...    APTMQLRSKCLVLPLLGLCIFANRAVTLQHTATLKSFDELRIECSRYLPPVDALNNVEDCSDRCLGLVGRFWNDSISRTVYSVSRFYQPDSCDQDNLDRTEQCLCETVQSLPRNASCQRASCSMQCYQDQFGELINQKPQFVPVSKL
AAEL003513/AaegOBP100/...    -----MGLKKWCIILPYLCSFAICVSALEHVATLKSFDEIRYECSQYLP----SSEDEDCSLRCLGLVGRFWNDTIGTPSNSVGRFYRPDSCDQCYLNRTEQCLRRTVLNLPRSAVCQRASNGLLCYKDQYGQLINRAPQFVPVVKL
AAEL003538/AaegOBP102/...    -----MGLKKWCIILPYLCSFAICVSALEHVATLKSFDEIRYECSQYLP----SSEDEDCSLRCLGLVGRFWNDTIGTPSNSVGRFYRPDSCDQCYLNRTEQCLRRTVLNLPGSAVCQRASNGLLCYKDQYGQLINRAPQFVPVVKL
AAEL003511/AaegOBP87/1...    ---MMSSGAFILSVVLSVSISVLQTSSLQHSATLKSFNEILSECSRYLP-----SNDEPCYDRCLGLVGRFWNDTIARPSVSVGRFYRPDPCDQNYVNRTQQCICDSVLPLPRKDVCLRASRGLQCYRNQYGRLIADEPLFVSVTPL
AAEL000821/AaegOBP6/12...    --------MKRLASLLLLFLTVRAEVSTQHCAVAKLPCTLHSECLQYLNS--NDDGPENCAYRCIALTARFWDDQKADVIRTISRFYLTDANDDFRNRTEQCLQETQETFPVTESCQRASCAFSCYNDQFGEVIAVRPSFIPFTAL
CPIJ008161/CquiOBP81         SVPFTHMTMKLFWRTLLVASALSCATGRRHNAKLKSLLAVDRECNQYST----SDSYGQCQTRCQGILGRFWDDSDVVPAVTLGRFYQPDCDDHNYLNRTEQCLGEVTDHVRRNDDCLRATCTVKCYTDQYGQLDAESPSYIPTTKL
AAEL010875/AaegOBP103/...    --MQVFAAIVKLVAITLGAIIASISCIEEHSASLKSILSSSAECNLYLPT---EALRQECGTRCVSLVNRIWNDTNGRLSDTIGRFYVEGPQDPCARNRTLQCLEQVTASIPLRNSCQLADASVNCYRNNYGQLDVKSPRFVAFSDV
AAEL010874/AaegOBP88/1...    --MQVFAAIVKLVAITLGAIIASISCIEEHSASLKSILSSTAECNLYLPT---EALRQECGTRCVSLVNRIWNDTNGRLSDTIGRFYVEGPQDPCARNRTLQCLEQVTASIPIRNSCKLADASVNCYRNNYGQLDVKSPRFVAFSDV
CPIJ008157/CquiOBP75         ----MKLLLVGCTFFAVVVRDVVHGDCHHHNAVFKSVLATDHECGIYLGD--RTCDGGDCRKRCRGLLTRYWNDSLGVPSATINRFYRPSVHNLEYHERTECCLSDVRKKVPLMEVCRRADASISCYNQHYGNLEAEIPAFIPFTDL
CPIJ008159/CquiOBP77         ------MKLILAVFAFAAIFGCIRADCRHHNAIFKSVLATDHECSIYLREQ-TTCDGEDCTVRCRGLLTRYWNDSVGVPSATINRFYRPSVLDLGYSERTGCCLREVIARVPQRAVCRRADATIGCFNQHYGHFRVKSPKFVPFTDL
CPIJ008158/CquiOBP76         ---------MKLLLTYLAALSAVFGDCSHHNAVFKSIHGADHECSLYLRE--GTCHSADCSLRCRGLLTRYWNDTVAVPSATINRFYLPSAHDYEYYERTECCLGGVKEKVPLEDVCRRAEASTDCYHRNYGQLQADTPTFVPFTDL
CPIJ008155/CquiOBP78         ----MKRLLAVLATILTLHSDVASSESGYHEAVFKSVRRADRECSIFLRD--GTCGKDDCSVRCRGLVTRFWNDSVGVPAVTIDRFYKPAHDDYEYYDLTEQCVGEVVAAVPCTDVCGRADSTVQCYNDQYGKLDKKNPKFVPFTKL
CPIJ008156/CquiOBP79         -------MKLLPAAVILTFTIDVAVSVGHHEAVFKSVRQARRECSVHLRD--GTAGRGECSTRCRGLVTRYWNDSVGFPAETIDRFYQPAADDGEYYDLTEQCLGEVTATIPVEDSCRRADCTVQCYGDQYGQLDGETAKFVPFTDL
CPIJ008160/CquiOBP80         ----------MQPLLTILTGSLLALSVEAHNAIFKSIHSVDRECVLYLRD--GTCGKGDCTVRCRGLVARFWKDTVGRPAATINRFYRPDPEDCGYYERTEQCLGEISAVIPAEDACRRANCSVSCYEDQYGRVDRYGVRFIPFTEL
CPIJ008154/CquiOBP82         ------------MLSTVLLGLLTVTQASNHLASFKSFDSTDRECSRYL------GHQHRCAVRCRGLLTRFLNRTSEFPTAVIERFYEPGEDDYLYRSRTVQCLRDVRAL----NSCEQARRSVQCYDGEFGTVDRDALRFVPFTDV
AAEL000835/AaegOBP97/1...    --------MFRIGLFFASFAVVSITAVDRHKIVYKSLQEAAVECGQY-------TIKGQCLGRCETLITRDWNDTTGM-SPAYSRFFHPDPVDECNLNRTQRCLQTKVYTVPRPRTCQRASESIQCYLDQFGQVNLTAPQFVRFTPL
AAEL000796/AaegOBP96/1...    --------MFRIGLLFVSFAVVSITAVDRHKIVYKSLQEAAVECGQY-------TIKGQCLGRCETLITGDWNDTTGM-SPAYSRFFHPDPVDECNLNRTQRCLQTKVYTVPRPRTCQRASESIQCYLDQFGQVNLTAPQFVRFTPL
AAEL000833/AaegOBP7/12...    -------MILHLRILFAVIALVSFCSCNQHKIVQKSLGKANDECVSY-------GSSEQCLARCVTLVTRDWNETVGL-SSVYDRFYQPDPEDLCNTNRTQRCLEALQSTVAPEDKCLRAAGSVQCYLDQYGQVDMATSRFVKSAPV
AAEL000831/AaegOBP85/1...    -------MSCLNLATLILAFFATSVICTRHKIVQKSLAGTDIECRQY-------DPPWNCAVRCQTLLMRDWVDSTGMQSP-YDRFFQPDSNDRCNTNRTQRCLLDKLSTLPRNKLCLRADISVQCFLNQSGQVIMDQPKFVAPSRL
AAEL000827/AaegOBP84/1...    -------MRRFKLASFILTLFATNVICSRHKIVQKSLAGTGVECQQY-------DPPWNCAVRCQTLLTRDWVDSTGMQSP-YDRFFQPDPNDQCYMNRTQRCLLDKLSTVPRNKLCLRADSSVQCFLNQSGQVIMDQPKFVAPSRL
AAEL004856/AaegOBP86/1...    -------MTIFNALLAILACFSLPTDALQHNAVYKSINSAGPECRTILT----RQSPLDCRLRCLSINTGDWDDCSGVP-RTYDRFYVQDPTDVGYQQRTQQCIANVSVSILRGDICAFSARSTECYDANYYDIVLDQLVFVPSKSL
AAEL010872/AaegOBP46/1...    -------MSGSISLIVLAVVALAGQVLGRHDATFKSFGSTSGECSRYLN----NDGNGECNIHCVGVIGHAWNETLAKFTQNYAGYFVPDPQDDCYQNRTERCLLQVDNAIPVYDKCTRASKLGQCYADQYGQLNAIQPQYVPMTDL
AAEL003315/AaegOBP16/1...    ----------MIVIVVLVALSSEVLGAGQHDAVFKSIESTGKECARYLN----NDGTGDCNTHCVGVIDHVWNDTVAMFTRNYERFFVPAPEDLCYQNRTQRCLSQVDQVVPVSDKCARARQLGQCYADQYGQLNASQLQYRPMTNL
```

Supplementary Figure 3c. Protein sequence alignments of OBPs that belongs to the different clusters of the *Atypical* subfamily. See additional legend for details.

Supplementary Figure 2c

Atypical : matype4 (average seq. identity : 33.5%) - continued from previous page



Additional legend to Figure 3c.

Shown are the alignments for the *Atypical* OBPs from the three mosquito genomes *Anopheles gambiae, Aedes aegypti* and *Culex quinquefasciatus*. These were obtained after aligning the sequences with a structure-based profile using ClustalX software. The global *Atypical* OBP alignment was splitted into different parts corresponding to the phylogenetic clusters that is established in this study (see Figures 4 and 5). The residues that are highly conserved (with 75% or more degree of conservation) are highlighted in the sequences and in the sequence logos above the alignment. The consensus sequence is also featured on top of each alignment. The average pairwise sequence identities within each cluster are indicated. These diagrams were generated by the Geneious software.

Supplementary Figure 3c. Protein sequence alignments of OBPs that belongs to the different clusters of the *Atypical* subfamily. See additional legend for details.