

Analiza filogenetyczna i tworzenie map genomowych, dla rzędu Hymenoptera (Błonkoskrzydłe). Paweł Grygielski, Michał Somala, Jan Fijor.

Zbiór danych:

Do stworzenia drzewa filogenetycznego wykorzystamy sekwencje genu COI następujących przedstawicieli Hymenoptera: *Allantus togatus*, *Apis dorsata* (pszczoła olbrzymia), *Apis mellifera* (pszczoła miodna), *Blasticotoma filiceti*, *Camponotus atrox*, *Dentathalia scutellariae*, *Lepisiota frauenfeldi*, *Tapinoma ibericum*, *Trichogramma pretiosum*, *Vespa simillima*, *Vespa velutina*. Jako outgrupę wybraliśmy *Drosophila hydei* należącą do Diptera. Zarówno Diptera, jak i analizowany przez nas rząd Hymenoptera należą do kladu Endopterygota (skrytoskrzydłe).

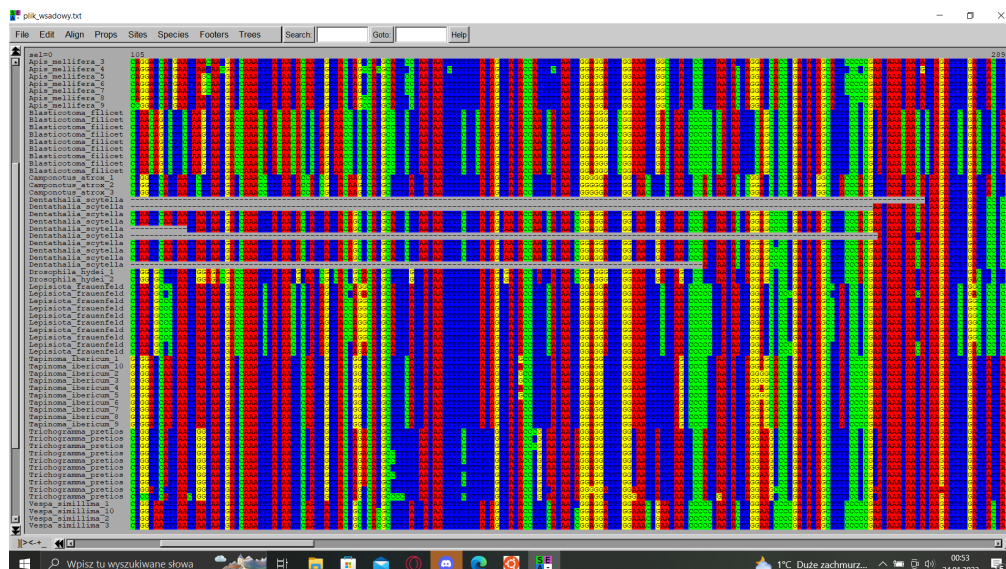
Do tworzenia map genomowych wykorzystamy 20 genomów mitochondrialnych. Wśród nich znajdują się wszystkie gatunki należące do Hymenoptera wymienione wcześniej, a także: *Anoplolepis gracilipes*, *Athalia arunachalensis*, *Athalia sikkimensis*, *Megabeleses liriodendrovorax*, *Megabeleses magnoliae*, *Neostromboceros nipponicus*, *Nesodiprion zhejiangensis*, *Runaria punctata*, *Trichogramma cacaeciae*. Genomem referencyjnym będzie *Apis mellifera*.

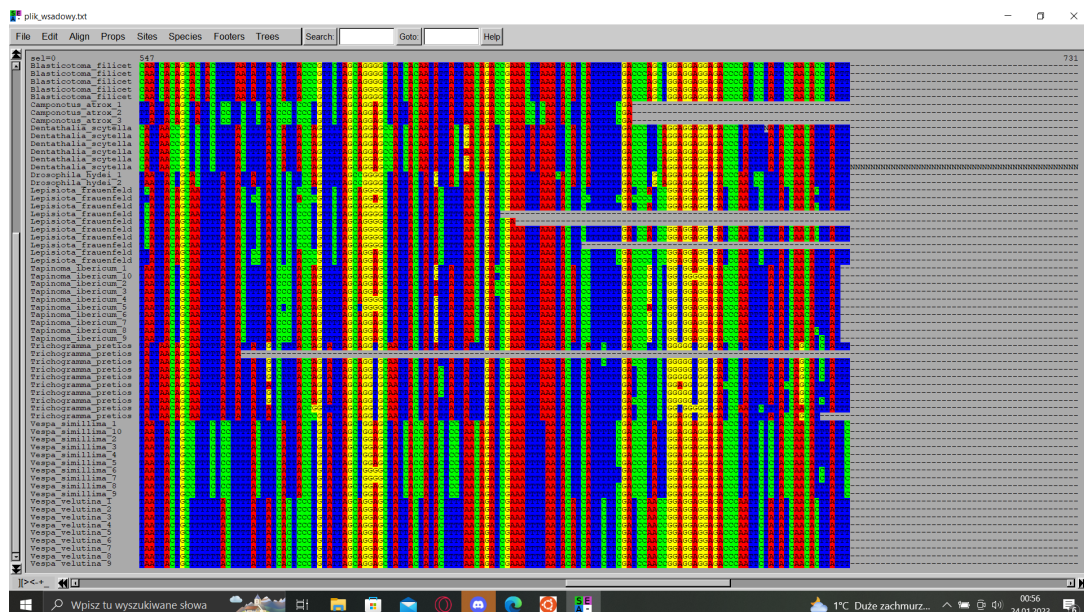
Po zebraniu sekwencji dostosowaliśmy je tak, aby nazwa gatunkowa znajdowała się na początku, tak by drzewo i mapa były czytelne.

Wszystkie sekwencje pochodzą z bazy NCBI.

Tworzenie drzewa filogenetycznego

Pierwszym krokiem do stworzenia drzewa filogenetycznego jest odpowiednie przycięcie sekwencji, w programie seaview.

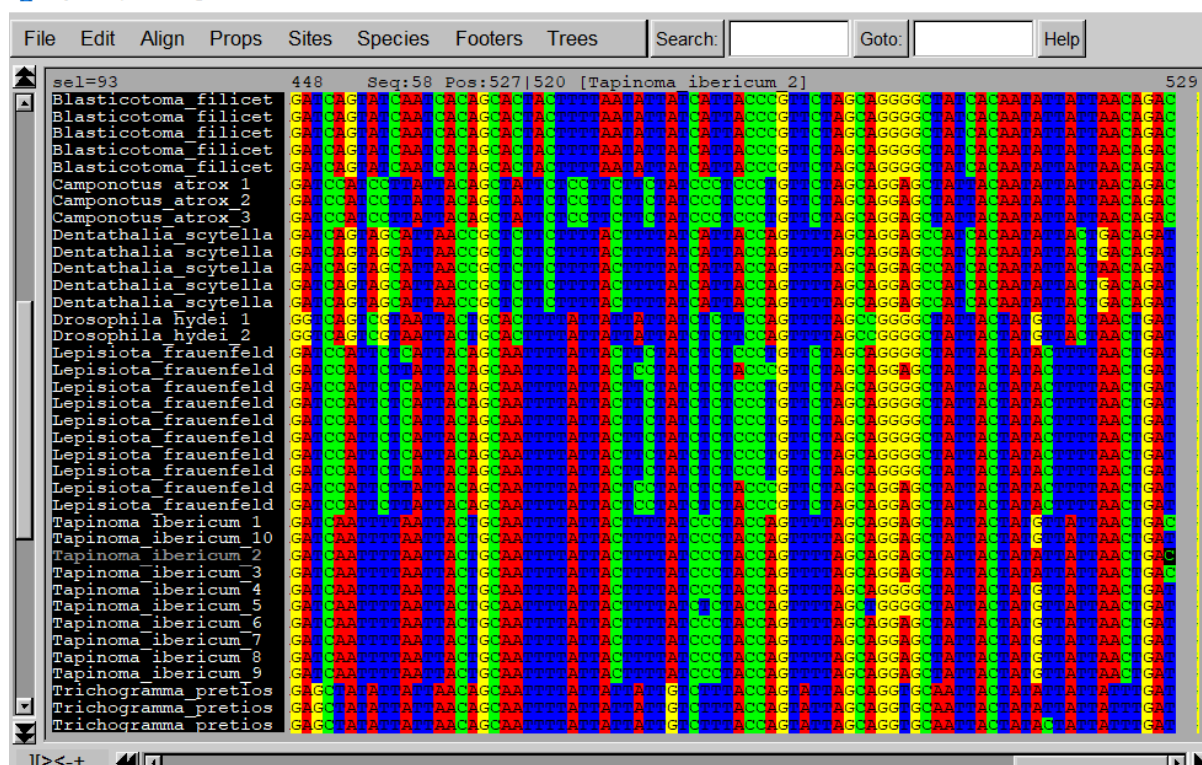




Aby przyrównanie nie było zbyt krótkie, wytniemy sekwencje, zaczynają się bardzo późno, lub kończą zbyt wcześnie. Ze 102 sekwencji (łącznie z outgrupą) pozostawiliśmy 93 i odpowiednio je przycięliśmy.

W celu stworzenia alignmentu skorzystamy z tego samego programu i zawartego w nim algorytmu muscle.

Hymenoptera_alignment



Otrzymaliśmy przyrównanie o długości 520 bp.

Aby wybrać odpowiednie modele substytucji, podzielimy modele substytucji na 1, 2 i 3 literę kodonu i sprawdzimy modele substytucji również dla nich, by były jak najlepiej dopasowane.

Dla wszystkich 4 plików wykonamy kalkulacje AIC, AICC oraz BIC

Wyniki kalkulacji dla całego genu:

Results										
Models AIC AICC BIC DT										
ID	Name	Partition	-lnL	p	AIC	deltaAIC	weight	cumtWeight	uDelta	
7	FBI+G	000000	5351.6344	188	11289.5174	277.1391	0.0	1	-	
9	FBI+I+G	000000	5344.3448	189	11279.8051	207.4271	0.0	1	-	
8	K80	010010	5405.5372	185	11318.8006	119.5442	0.0	1	-	
10	K80+I	010010	5544.2459	186	11665.0919	651.7136	0.0	1	-	
11	K80+G	010010	5514.532	186	11601.6639	591.2857	0.0	1	-	
12	K80+I+G	010010	5507.6692	187	11596.7485	584.3701	0.0	1	-	
13	HKY	010010	5627.7601	188	11841.7467	826.3904	0.0	1	-	
14	HKY+I	010010	5311.1936	189	11213.5029	201.1246	0.0	1	-	
15	HKY+G	010010	5215.7002	189	11022.5161	10.1379	0.0059	0.994	-	
16	HKY+I+G	010010	5208.1832	190	11007.2293	0.0	0.9836	0.9836	-	
17	SVM	012345	5609.5849	189	11810.2855	797.9072	0.0	1	-	
18	SVM+I	012345	5358.6147	190	11313.2412	300.8629	0.0	1	-	
19	SVM+G	012345	5376.0571	190	11348.1261	335.7479	0.0	1	-	
20	SVM+I+G	012345	5318.8373	191	11051.8746	0.0	1	-	-	
21	GTR	012345	5562.5737	192	11731.0396	718.6613	0.0	1	-	
22	GTR+I	012345	5276.5606	193	11161.9981	151.6199	0.0	1	-	
23	GTR+G	012345	5205.8516	193	11022.58	10.2017	0.0059	0.994	-	
24	GTR+I+G	012345	5201.2518	194	11008.5947	6.2164	0.0043	0.9957	-	
Decimal numbers are rounded. Click on column headers to sort data in ascending or descending order (+Shift)										
24 stycznia 2023										

Results										
Models AIC AICC BIC DT										
ID	Name	Partition	-lnL	p	AIC	deltaAIC	weight	cumtWeight	uDelta	
7	FBI+G	000000	5351.6344	188	11289.5174	277.1391	0.0	1	-	
9	FBI+I+G	000000	5344.3448	189	11279.8051	207.4271	0.0	1	-	
8	K80	010010	5405.5372	185	11318.8006	119.5442	0.0	1	-	
10	K80+I	010010	5544.2459	186	11665.0919	651.7136	0.0	1	-	
11	K80+G	010010	5514.532	186	11601.6639	591.2857	0.0	1	-	
12	K80+I+G	010010	5507.6692	187	11596.7485	584.3701	0.0	1	-	
13	HKY	010010	5627.7601	188	11841.7467	826.3904	0.0	1	-	
14	HKY+I	010010	5311.1936	189	11213.5029	201.1246	0.0	1	-	
15	HKY+G	010010	5215.7002	189	11022.5161	10.1379	0.0059	0.994	-	
16	HKY+I+G	010010	5208.1832	190	11007.2293	0.0	0.9836	0.9836	-	
17	SVM	012345	5609.5849	189	11810.2855	797.9072	0.0	1	-	
18	SVM+I	012345	5358.6147	190	11313.2412	300.8629	0.0	1	-	
19	SVM+G	012345	5376.0571	190	11348.1261	335.7479	0.0	1	-	
20	SVM+I+G	012345	5318.8373	191	11051.8746	0.0	1	-	-	
21	GTR	012345	5562.5737	192	11731.0396	718.6613	0.0	1	-	
22	GTR+I	012345	5276.5606	193	11161.9981	151.6199	0.0	1	-	
23	GTR+G	012345	5205.8516	193	11022.58	10.2017	0.0059	0.994	-	
24	GTR+I+G	012345	5201.2518	194	11008.5947	6.2164	0.0043	0.9957	-	
Decimal numbers are rounded. Click on column headers to sort data in ascending or descending order (+Shift)										
24 stycznia 2023										

Results										
Models AIC AICC BIC DT										
ID	Name	Partition	-lnL	p	BIC	deltaBIC	weight	cumtWeight	uDelta	
1	JC	000000	5838.516	184	12630.197	1223.6025	0.0	1	-	
2	JC+I	000000	5581.6408	185	12322.7137	715.5793	0.0	1	-	
3	JC+G	000000	5556.8597	185	12273.1516	666.6171	0.0	1	-	
4	JC+I+G	000000	5546.21	186	12258.1193	650.9849	0.0	1	-	
5	FBI	000000	5693.7697	187	12559.5458	952.4114	0.0	1	-	
6	FBI+I	000000	5409.0497	188	11996.313	389.1968	0.0	1	-	
7	FBI+G	000000	5351.6344	188	11881.5026	274.3681	0.0	1	-	
8	FBI+I+G	000000	5344.3448	189	11873.1905	266.056	0.0	1	-	
9	K80	010010	5405.5372	185	12776.3603	1163.372	0.0	1	-	
10	K80+I	010010	5544.2459	186	12254.1912	647.6567	0.0	1	-	
11	K80+G	010010	5514.532	186	12194.7632	587.6288	0.0	1	-	
12	K80+I+G	010010	5507.6692	187	12187.305	580.1705	0.0	1	-	
13	HKY	010010	5627.7601	188	12433.7539	826.6194	0.0	1	-	
14	HKY+I	010010	5311.1936	189	11886.8881	199.7536	0.0	1	-	
15	HKY+G	010010	5215.7002	189	11615.9013	8.7669	0.0123	0.996	-	
16	HKY+I+G	010010	5208.1832	190	11607.1345	0.0	0.9836	0.984	-	
17	SVM	012345	5609.5849	189	12403.6707	796.5362	0.0	1	-	
18	SVM+I	012345	5358.6147	190	11907.8974	280.8629	0.0	1	-	
19	SVM+G	012345	5376.0571	190	12000.8629	393.0655	0.0	1	-	
20	SVM+I+G	012345	5318.8373	191	11601.8746	0.0	1	-	-	
Decimal numbers are rounded. Click on column headers to sort data in ascending or descending order (+Shift)										
24 stycznia 2023										

Po kalkulacjach dla wszystkich plików otrzymujemy następujące modele.

	jModelTest			
	AIC	AICc	BIC	WYBÓR
Cały gen	GTR+I+G	HKY+I+G	HKY+I+G	HKY+I+G
1st	GTR+G	HKY+G	HKY+G	HKY+G
2nd	GTR+I+G	HKY+I+G	HKY+I+G	HKY+I+G
3rd	GTR+G	HKY+G	HKY+I+G	HKY+I+G

Znając optymalne modele substytucji dla całego genu i kodonów, możemy przejść do tworzenia drzew(z partycjami i bez partycji). Wykonamy je w programie MrBayes. W tym celu tworzymy chartsety, a następnie partycje, dla których zakodujemy odpowiednie modele.

Tworzenie drzewa bez partycji:

1. kodowanie modeli dla kolejnych partycji:

```
MrBayes > lset applyto=(2,3) nst=2 rates=invgamma

Setting Nst to 2 for partition 2
Setting Nst to 2 for partition 3
Setting Rates to Invgamma for partition 2
Setting Rates to Invgamma for partition 3
Successfully set likelihood model parameters to
partitions 2 and 3 (if applicable)

MrBayes > lset applyto=(1) nst=2 rates=gamma

Setting Nst to 2 for partition 1
Setting Rates to Gamma for partition 1
Successfully set likelihood model parameters to
partition 1 (if applicable)
```

Po przypisaniu odpowiednich modeli zrobić konkretną ilość kroków w łańcuchach Markowa.

```
Average standard deviation of split frequencies: 0.010227

1216000 -- (-5397.976) [-5351.858] (-5401.544) (-5361.569) * [-5358.061] (-5410.036) (-5434.620) (-5399.749) -- 1:45:12
1217000 -- (-5387.392) [-5353.661] (-5386.377) (-5396.150) * (-5370.828) (-5407.097) (-5431.169) [-5392.209] -- 1:45:05
1218000 -- (-5394.953) [-5376.382] (-5372.138) (-5397.396) * [-5378.289] (-5425.211) (-5450.652) (-5385.940) -- 1:44:58
1219000 -- (-5410.935) (-5393.203) (-5381.164) [-5370.851] * (-5378.751) [-5388.909] (-5415.594) (-5412.830) -- 1:44:52
1220000 -- (-5403.730) (-5366.473) [-5352.801] (-5398.753) * [-5372.931] (-5396.569) (-5417.101) (-5400.313) -- 1:44:45

Average standard deviation of split frequencies: 0.009987

1221000 -- (-5380.256) (-5382.064) [-5352.756] (-5420.044) * (-5371.053) (-5390.428) [-5372.003] (-5393.415) -- 1:44:37
1222000 -- (-5397.947) [-5386.646] (-5362.157) (-5425.388) * (-5398.260) (-5399.021) [-5371.344] (-5385.265) -- 1:44:31
1223000 -- (-5396.495) [-5373.896] (-5354.150) (-5406.880) * (-5389.546) (-5417.920) [-5376.579] (-5388.866) -- 1:44:24
1224000 -- (-5398.296) [-5356.832] (-5349.109) (-5403.888) * (-5402.290) (-5406.585) [-5362.068] (-5396.027) -- 1:44:17
1225000 -- (-5398.244) [-5359.192] (-5378.763) (-5389.629) * (-5380.989) (-5396.436) [-5337.571] (-5405.351) -- 1:44:10

Average standard deviation of split frequencies: 0.009734

1226000 -- (-5377.971) [-5361.987] (-5371.612) (-5387.245) * (-5407.535) (-5388.172) [-5346.917] (-5388.275) -- 1:44:03
1227000 -- (-5367.379) [-5357.685] (-5412.056) (-5375.191) * [-5362.407] (-5415.801) (-5348.413) (-5388.829) -- 1:43:56
1228000 -- (-5381.107) (-5388.360) (-5379.137) [-5369.336] * (-5406.870) (-5391.079) [-5365.052] (-5384.113) -- 1:43:49
1229000 -- (-5412.799) (-5383.521) (-5372.456) [-5380.866] * (-5400.460) (-5395.015) (-5390.242) [-5363.874] -- 1:43:42
1230000 -- (-5394.597) (-5395.966) (-5403.989) [-5367.490] * (-5382.593) (-5406.544) (-5389.552) [-5370.694] -- 1:43:36
```

Dojście do konwergencji(wartość odchylenia spadła poniżej 0,01) zajęło nam 1230000 kroków, musimy kontynuować analizę jeszcze przez 3690000(trzykrotność kroków, którą musieliśmy wykonać by dotrzeć do konwergencji), aby otrzymać poprawne drzewo

```
Average standard deviation of split frequencies: 0.005340

4916000 -- [-5372.921] (-5411.713) (-5424.495) (-5371.873) * (-5399.607) [-5375.853] (-5379.201) (-5420.152) -- 0:00:26
4917000 -- (-5366.184) (-5428.702) (-5434.972) [-5370.848] * (-5432.546) [-5381.397] (-5375.869) (-5423.518) -- 0:00:20
4918000 -- (-5368.598) (-5415.088) (-5440.827) [-5374.450] * (-5423.019) (-5393.372) [-5369.263] (-5386.756) -- 0:00:13
4919000 -- (-5338.851) (-5432.426) (-5428.300) [-5356.088] * (-5400.885) (-5407.597) [-5361.739] (-5393.026) -- 0:00:06
4920000 -- (-5365.251) (-5405.279) (-5399.403) [-5340.568] * (-5383.909) (-5405.147) [-5374.044] (-5397.445) -- 0:00:00

Average standard deviation of split frequencies: 0.005293

Continue with analysis? (yes/no): sump
Enter yes or no: no

Analysis completed in 9 hours 13 mins 25 seconds
Analysis used 33206.11 seconds of CPU time
Likelihood of best state for "cold" chain of run 1 was -5319.55
Likelihood of best state for "cold" chain of run 2 was -5324.01

Acceptance rates for the moves in the "cold" chain of run 1:
With prob. (last 100) chain accepted proposals by move
25.7 % ( 23 %) Dirichlet(Tratio{all})
24.7 % ( 27 %) Dirichlet(Pi{all})
25.6 % ( 32 %) Slider(Pi{all})
26.0 % ( 20 %) Multiplier(Alpha{all})
29.9 % ( 32 %) Slider(Pinvar{2,3})
42.7 % ( 39 %) ExtSPR(Tau{all},V{all})
29.7 % ( 37 %) ExtTBR(Tau{all},V{all})
44.1 % ( 49 %) MNI(Tau{all},V{all})
```

Dla całego genu nie tworzymy charsetów i wybieramy jeden model dla całości, tym modelem będzie HKY+I+G

```
MrBayes > lset nst=2 rates=invgamma

Setting Nst to 2
Setting Rates to Invgamma
Successfully set likelihood model parameters

MrBayes > _
```

```
Average standard deviation of split frequencies: 0.009957
1066000 -- (-5378.426) (-5371.208) (-5409.300) (-5409.770) * (-5427.775) (-5381.135) [-5364.696] (-5387.044) -- 0:38:
47 1067000 -- (-5410.269) [-5375.109] (-5377.340) (-5445.059) * (-5457.035) (-5376.653) [-5376.736] (-5384.384) -- 0:38:
42 1068000 -- (-5416.110) (-5391.243) [-5371.678] (-5416.117) * (-5427.087) (-5356.543) (-5384.255) [-5367.432] -- 0:38:
37 1069000 -- (-5428.508) (-5391.491) [-5372.029] (-5392.761) * (-5415.798) (-5367.045) (-5388.670) [-5349.936] -- 0:38:
31 1070000 -- (-5436.948) (-5392.081) (-5397.859) [-5382.330] * (-5409.702) (-5378.739) (-5418.372) [-5353.533] -- 0:38:
26

Average standard deviation of split frequencies: 0.009908
1071000 -- (-5435.710) (-5388.730) (-5396.007) [-5379.550] * (-5399.166) (-5383.724) (-5405.526) [-5346.467] -- 0:38:
21 1072000 -- (-5411.587) (-5398.880) [-5375.312] (-5395.154) * (-5425.977) [-5375.808] (-5409.061) (-5355.515) -- 0:38:
15 1073000 -- (-5386.075) (-5378.589) [-5380.042] (-5405.607) * (-5414.970) (-5390.198) (-5385.846) [-5374.178] -- 0:38:
10 1074000 -- (-5379.144) (-5367.561) (-5365.184) [-5386.054] * (-5404.264) (-5395.051) (-5368.465) [-5360.814] -- 0:38:
05 1075000 -- (-5388.311) (-5373.803) [-5354.261] (-5411.385) * (-5419.432) (-5407.901) (-5348.757) [-5356.153] -- 0:37:
59
```

Tym razem dojście do konwergencji zajęło 1070000 kroków, zatem musimy wykonać jeszcze 3210000.

```
Average standard deviation of split frequencies: 0.005226
4276000 -- (-5448.061) [-5344.843] (-5389.842) (-5395.503) * (-5403.203) (-5396.575) (-5409.348) [-5376.041] -- 0:00:22
4277000 -- (-5430.292) [-5351.682] (-5406.926) (-5409.923) * (-5398.692) (-5406.953) (-5390.262) [-5361.500] -- 0:00:16
4278000 -- (-5397.307) [-5366.116] (-5420.041) (-5434.592) * (-5418.237) (-5403.412) (-5396.523) [-5371.952] -- 0:00:11
4279000 -- (-5397.064) [-5384.933] (-5436.843) (-5391.444) * (-5393.515) (-5392.884) (-5400.037) [-5387.246] -- 0:00:05
4280000 -- (-5396.492) (-5410.213) [-5379.622] (-5379.564) * [-5382.193] (-5406.891) (-5401.599) (-5390.652) -- 0:00:00

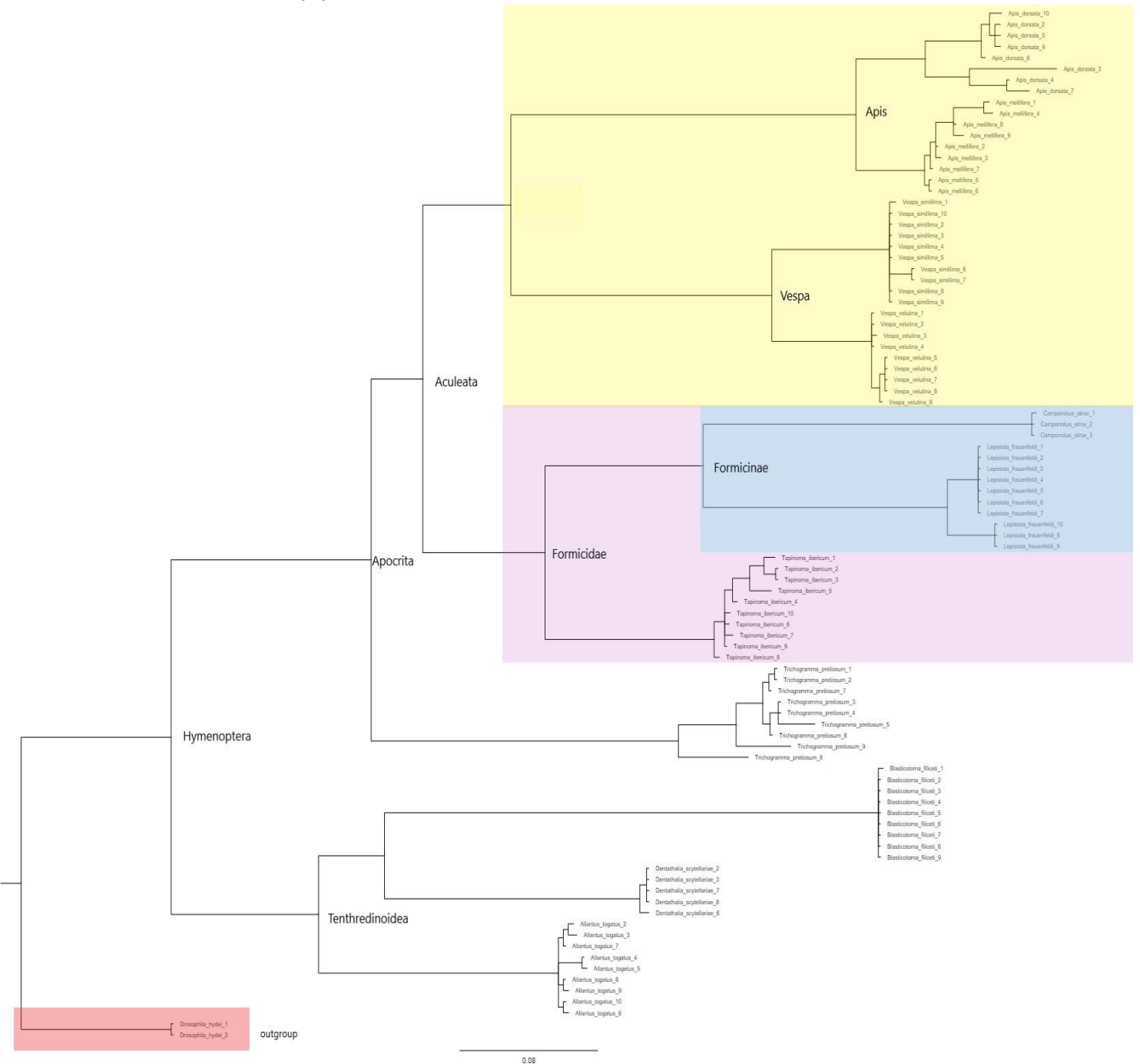
Average standard deviation of split frequencies: 0.005289

Continue with analysis? (yes/no): no

Analysis completed in 6 hours 32 mins 58 seconds
Analysis used 23578.64 seconds of CPU time
Likelihood of best state for "cold" chain of run 1 was -5321.08
Likelihood of best state for "cold" chain of run 2 was -5322.06
```

Oba drzewa ukorzenimy i dostosujemy w programie FigTree, a także dodamy taksony.

Drzewo bez partycji:

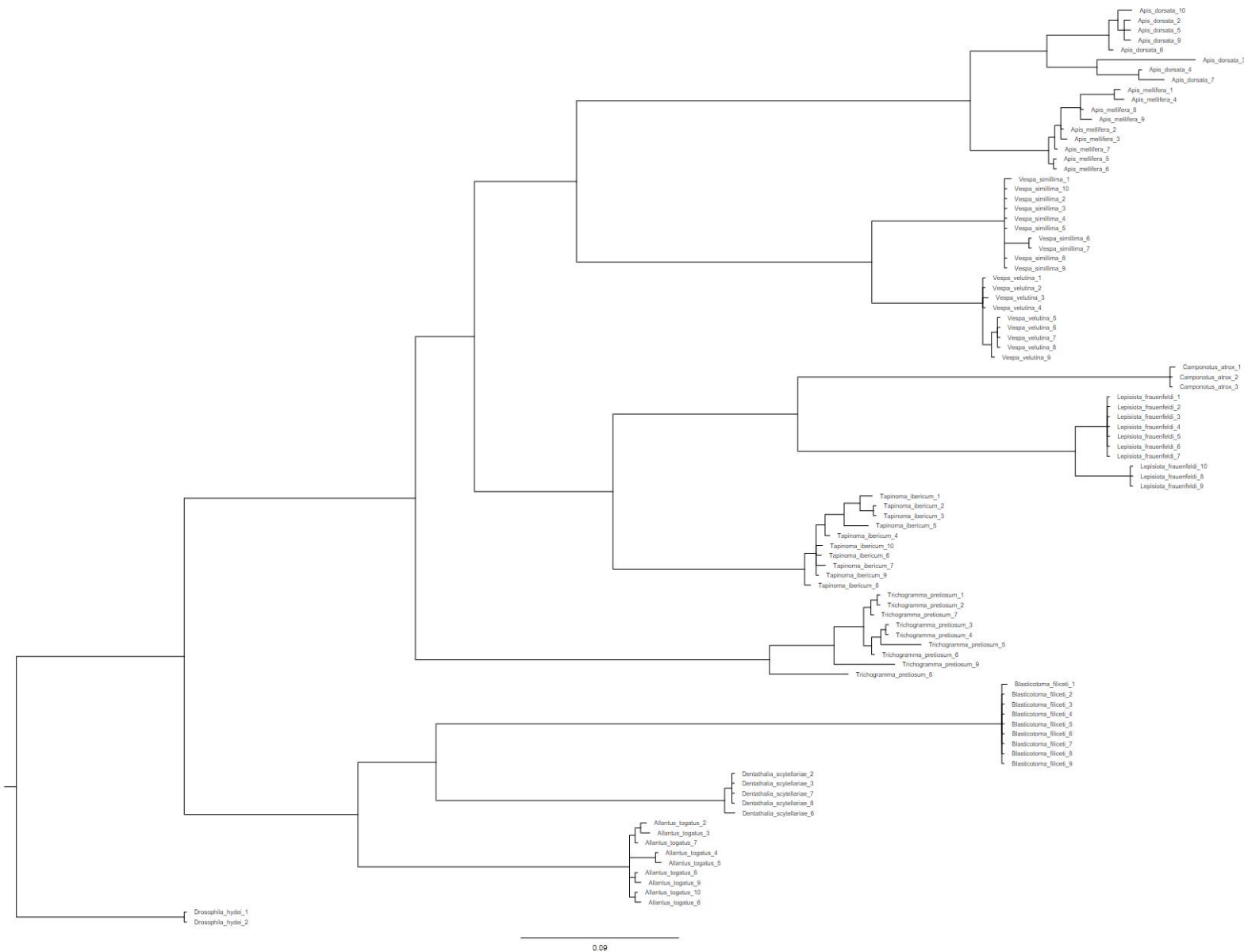


Różnice między gatunkami wynikać mogą z faktu, że sekwencje były różnej długości i rozpoczynały się w różnych miejscach genu, a także, w przypadku niektórych gatunków, np. *Apis mellifera*, z możliwej rozbieżności podgatunków.

Dobrze dostrzegalne są podobieństwa w przypadku mrówek, *Lepisiota* spp. i *Camponotus* spp. należą do Formicinae, natomiast *Tapinoma* spp. należą do innej podrodziny: Dolichoderinae i z Formicinae łączy je dopiero rodzina - Formicidae. Ta rodzina, należy z kolei do podrzędu Aculeata (do którego należą także Apidae i Vespidae). Wszystkie te taksony należą do Apocrita (wspólny przodek z *Trichogramma*) i dalej do Hymenoptera, do którego należy również Tenthredinoidea - nadrodzina, do której należą pozostałe gatunki z naszej analizy. Outgrupa zaznaczona została na czerwono, nie należy do Hymenoptera, a

do Diptera. Oba te rzędy należą do nadrzędu Endopterygota, co można zauważyć na drzewie.

Drzewo z partycjami wygląda niemal identycznie, więc wszystkie przedstawione wyżej informacje dotyczą także jego:



Identyczny wygląd drzewa najprawdopodobniej wynika z podobieństw modeli substytucji. Dla całego genu było to HKY+I+G, po podzieleniu ze względu na litery kodonu, dla 2 plików użyliśmy modelu HKY+I+G, a dla jednego HKY+G, podobieństwo nie powinno więc zaskakiwać

Tworzenie mapy genomowej

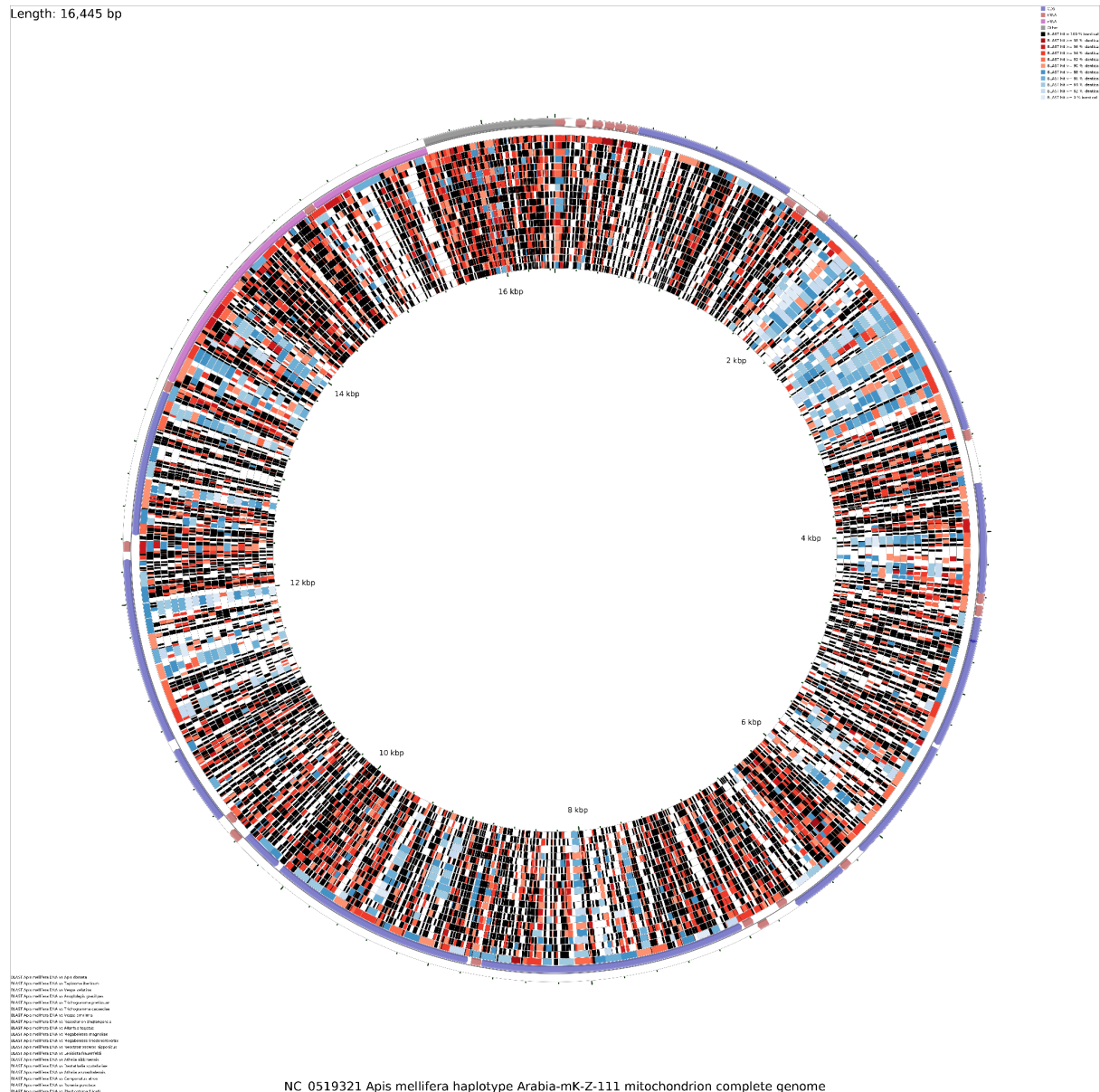
Do stworzenia mapy genomowej wykorzystamy sekwencje genomów mitochondrialnych przedstawicieli Hymenoptera, o których wspomnieliśmy na początku.

Tworzenie pliku GFF

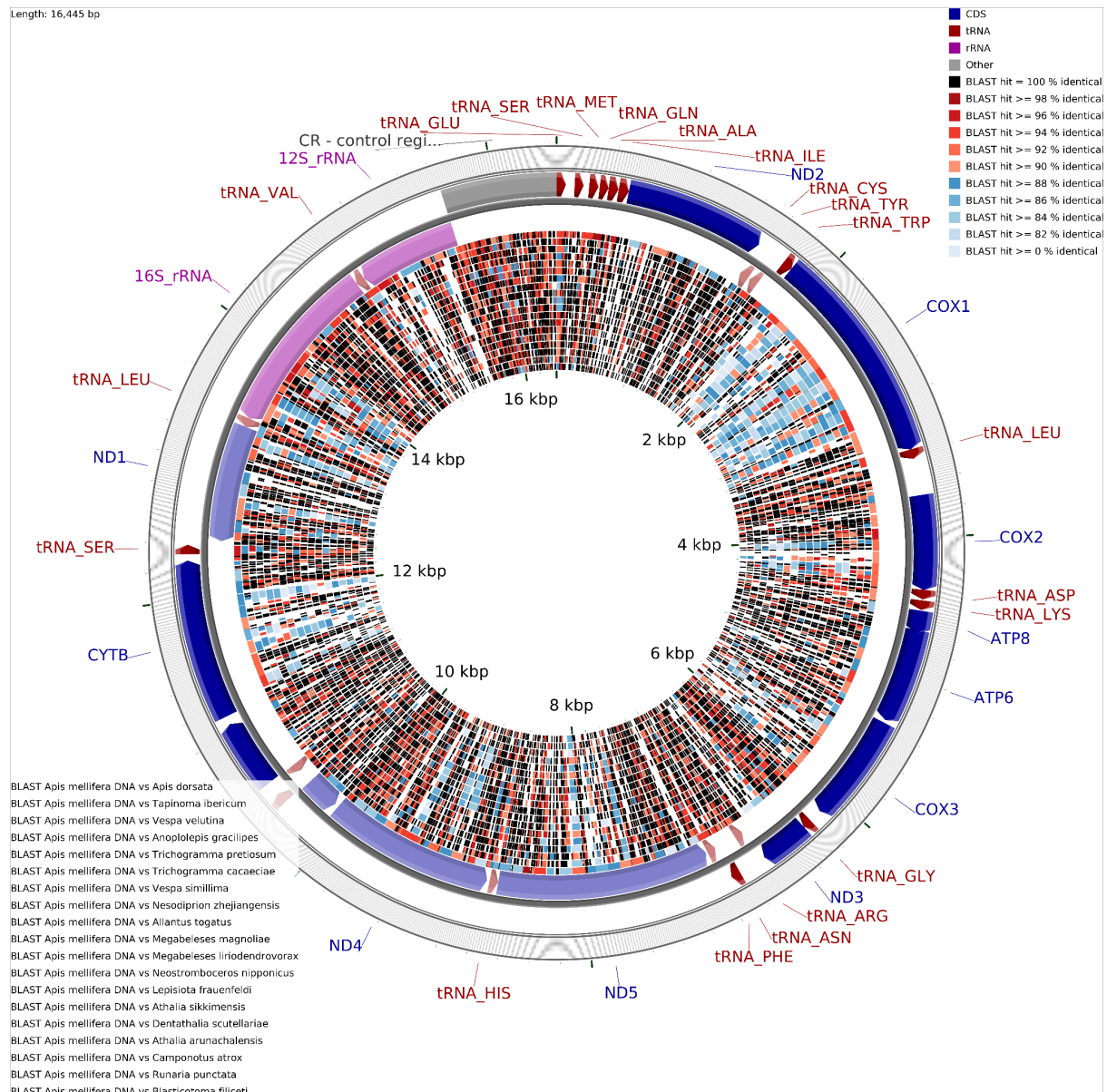
Do stworzenia mapy genomowej będziemy potrzebować łącznie 20 genomów, z których jeden zostanie genomem referencyjnym. W naszym przypadku będzie to genom *Apis mellifera*. Z tego samego genomu musimy stworzyć plik GFF, zrobimy to w arkuszu excel, by następnie wkleić informacje do pliku tekstowego:

	A	B	C	D	E	F	G	H
1	seqname	source	feature	start	end	score	strand	frame
2	tRNA_GLU	EMBL	tRNA	1	66	.	+	.
3	tRNA_SER	EMBL	tRNA	129	192	.	+	.
4	tRNA_MET	EMBL	tRNA	232	297	.	+	.
5	tRNA_GLN	EMBL	tRNA	307	369	.	+	.
6	tRNA_ALA	EMBL	tRNA	371	440	.	+	.
7	tRNA_ILE	EMBL	tRNA	443	515	.	+	.
8	ND2	EMBL	CDS	516	1517	.	+	.
9	tRNA_CYS	EMBL	tRNA	1517	1589	.	-	.
10	tRNA_TYR	EMBL	tRNA	1606	1673	.	-	.
11	tRNA_TRP	EMBL	tRNA	1734	1805	.	+	.
12	COX1	EMBL	CDS	1806	3366	.	+	.
13	tRNA_LEU	EMBL	tRNA	3367	3436	.	+	.
14	COX2	EMBL	CDS	3700	4377	.	+	.
15	tRNA_ASP	EMBL	tRNA	4376	4444	.	+	.
16	tRNA_LYS	EMBL	tRNA	4452	4520	.	+	.
17	ATP8	EMBL	CDS	4526	4684	.	+	.
18	ATP6	EMBL	CDS	4666	5346	.	+	.
19	COX3	EMBL	CDS	5367	6146	.	+	.
20	tRNA_GLY	EMBL	tRNA	6240	6305	.	+	.
21	ND3	EMBL	CDS	6306	6659	.	+	.
22	tRNA_ARG	EMBL	tRNA	6693	6759	.	-	.
23	tRNA_ASN	EMBL	tRNA	6862	6930	.	+	.
24	tRNA_PHE	EMBL	tRNA	6937	7004	.	-	.
25	ND5	EMBL	CDS	7017	8681	.	-	.
26	tRNA_HIS	EMBL	tRNA	8682	8749	.	-	.
27	ND4	EMBL	CDS	8770	10080	.	-	.
28	ND4L	EMBL	CDS	10111	10374	.	-	.
29	tRNA_THR	EMBL	tRNA	10387	10466	.	+	.
30	tRNA_PRO	EMBL	tRNA	10487	10556	.	-	.
31	ND6	EMBL	CDS	10565	11068	.	+	.
32	CYTB	EMBL	CDS	11128	12279	.	+	.
33	tRNA_SER	EMBL	tRNA	12325	12391	.	+	.
34	ND1	EMBL	CDS	12426	13343	.	-	.
35	tRNA_LEU	EMBL	tRNA	13344	13414	.	-	.
36	16S_rRNA	EMBL	rRNA	13415	14775	.	-	.
37	tRNA_VAL	EMBL	tRNA	14776	14846	.	-	.
38	12S_rRNA	EMBL	rRNA	14847	15634	.	-	.
39	CR - control region 1	EMBL	other	15635	16445	.	+	.

Length: 16,445 bp



Mapa jest mało czytelna i wymaga edycji, którą możemy przeprowadzić w pliku xml, a następnie narysować mapę jeszcze raz. W efekcie otrzymujemy znacznie bardziej czytelną mapę, widoczną na następnej stronie.



Na mapce widzimy, że najbardziej konserwatywnymi regionami są: region kontrolny, w większości przypadków geny kodujące aminokwasy, np. tRNA_GLY, tRNA ASN, tRNA_PHE, tRNA LEU itd. Silnie konserwatywny jest również obszar od 14kbp do 15kbp, oraz część genu ND4 (9,5-10,1kbp)

Zaskakujący jest fakt, że mało konserwatywnym obszarem wydaje się być gen COX1, dla którego tworzyliśmy drzewa, gdyż jest to gen opisywany jako silnie konserwatywny. Wrażenie takie, odnosimy ze względu na jasne kolory w tym obszarze, choć wniosek nie jest oczywisty, gdyż obszar ten ma stosunkowo mało miejsc z 0% identity, a trzeba zauważyć, że najniższe identity poza 0% to 82%.

Najmniej konserwatywne obszary to część genu CYTB (od 11,2kbp do 12kbp), oraz krótkie fragmenty genu 12S_rRNA (15-15,2 kbp, oraz 15,4-15,5 kbp).

Łatwo dostrzegalna jest także stosunkowo wysoka identyczność genomu *Apis dorsata* do genomu referencyjnego (*A. mellifera*). *Apis dorsata* jest bliżej spokrewniona z *A. mellifera*, niż jakikolwiek inny gatunek obecny na mapie, a skutkiem tego pokrewieństwa jest przynależność obu gatunków do tego samego rodzaju.