

**Dodatak radu:**

Naslov: Validacija R-P/F metode za analizu sličnosti sekvenci bez poravnanja na proteinskim sekvencama

Title: Validation of alignment-free R-P/F sequence similarity method on protein sequences

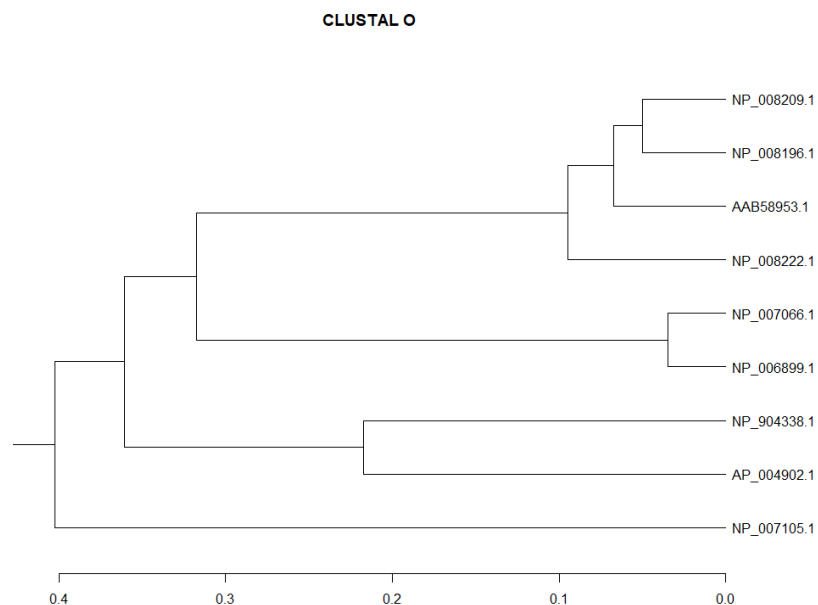
autor: Jasmina Jovanovic, master matematičar

Matematički fakultet, Univerzitet u Beogradu, 11000 Beograd, Srbija

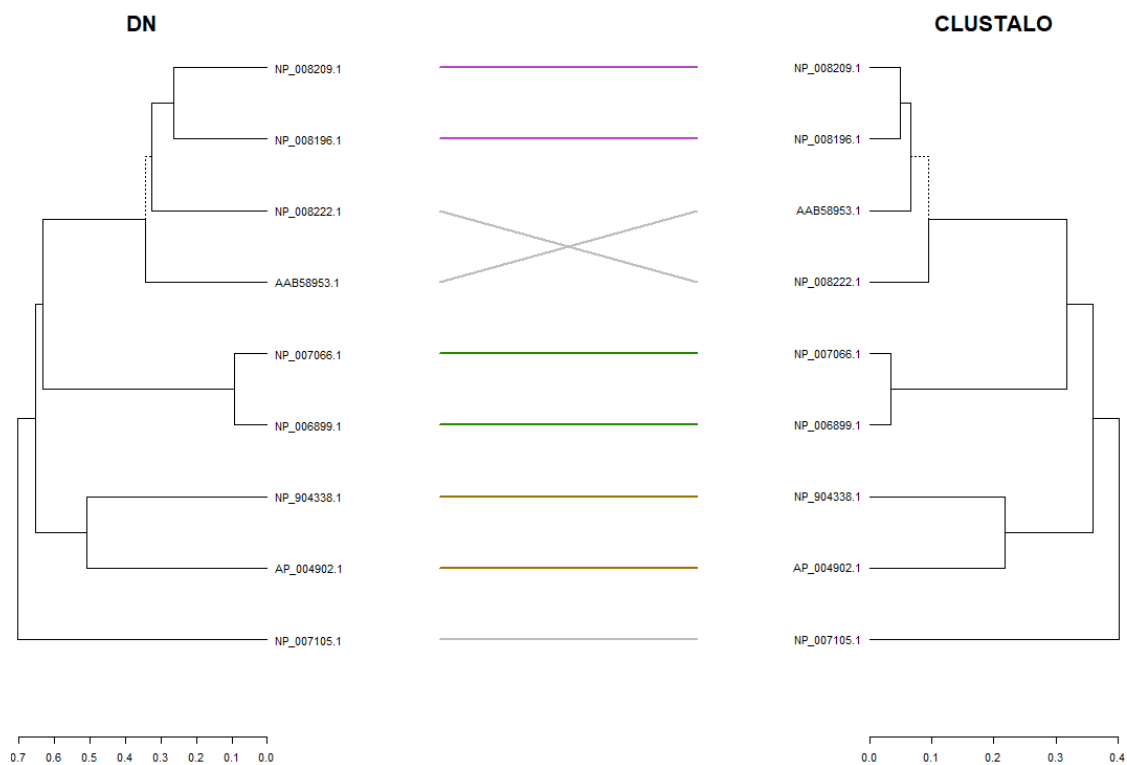
kontakt: [jasmina.dragoljevic@gmail.com](mailto:jasmina.dragoljevic@gmail.com)

Dodatna tabela 1: Sažete informacije o 9 MT-ND5 proteinskih sekvenci

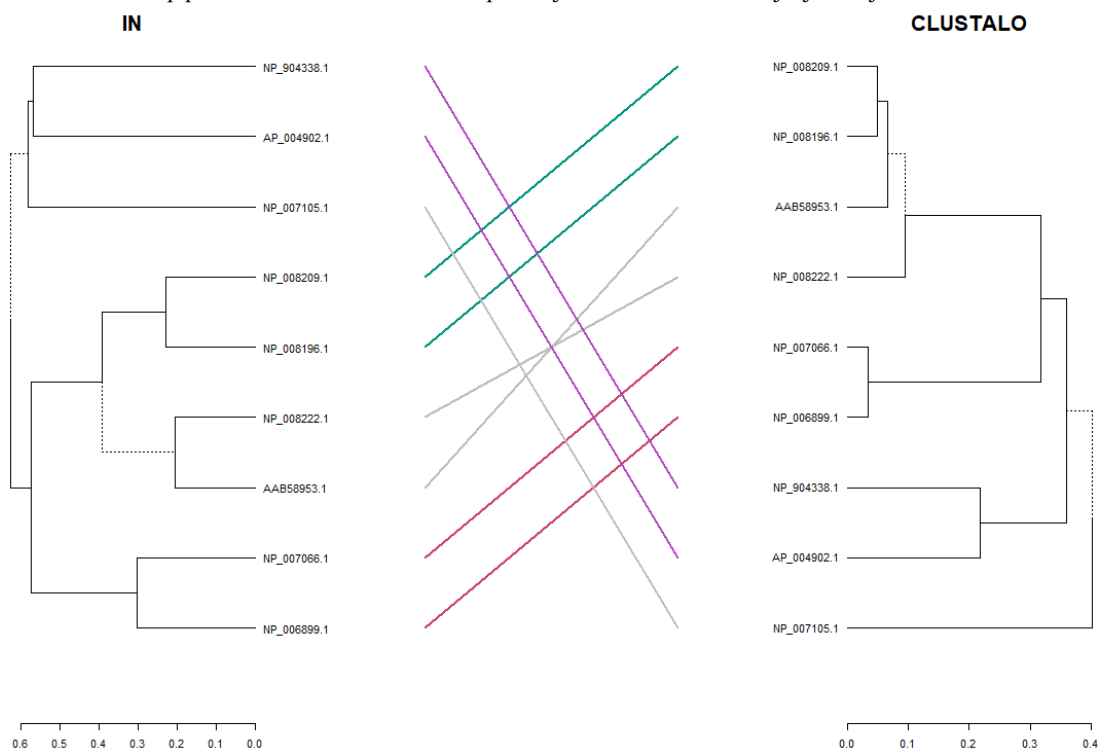
ID	Vrsta	Dužina proteinske sekvence
AAB58953.1	(primates) human	603
NP_008222.1	(primates) western gorilla	603
NP_008196.1	(primates) chimpanzee	603
NP_008209.1	(primates) pygmy chimpanzee	603
NP_006899.1	(whales & dolphins) fin whale	606
NP_007066.1	(whales & dolphins) blue whale	606
AP_004902.1	(rodents) Norway rat	610
NP_904338.1	(rodents) house mouse	607
NP_007105.1	(marsupials) North American opossum	602



Dodatna Figura 1: Dendrogramski prikaz sekvenci proteina NADH dehidrogenaze 5 (MT-ND5) konstruisano UPGMA metodom primenjenom na matricu rastojanja dobijenu iz ClustalΩ.



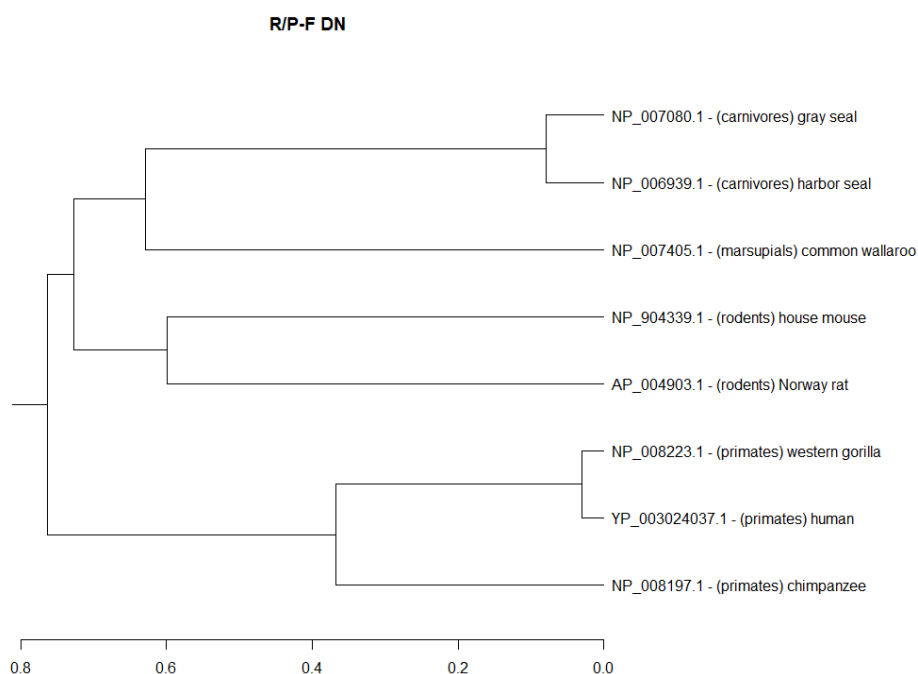
Dodatna Figura 2: Komparativni prikaz dendrograma sekvenci proteina NADH dehidrogenaze 5 (MT-ND5) konstruisano R-P/F metodom za DN tip ponovaka i UPGMA metodom primenjenom na matricu rastojanja dobijenu iz ClustalΩ.



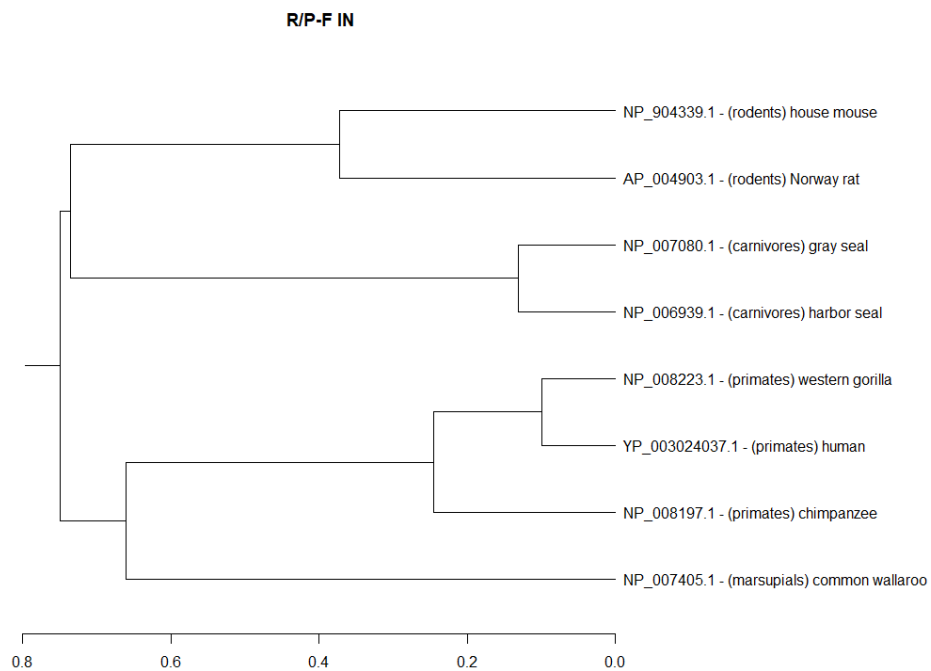
Dodatna Figura 3: Komparativni prikaz dendrograma sekvenci proteina NADH dehidrogenaze 5 (MT-ND5) konstruisano R-P/F metodom za IN tip ponovaka i UPGMA metodom primenjenom na matricu rastojanja dobijenu iz ClustalΩ.

Dodatna tabela 2: Sažete informacije o 8 MT-ND6 proteinskih sekvenci

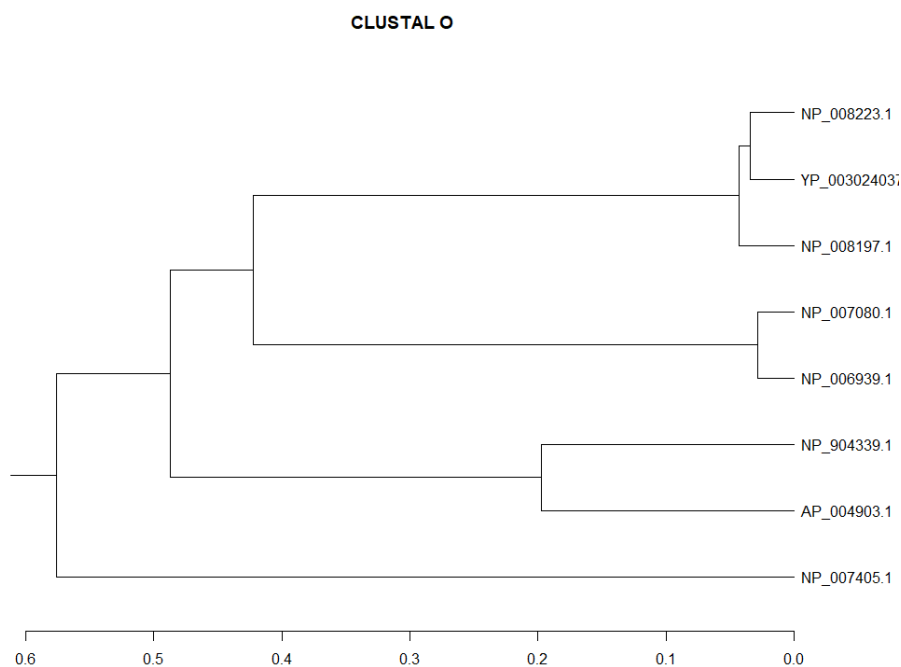
ID	Vrsta	Dužina proteinske sekvence
YP_003024037.1	(primates) human	174
NP_008223.1	(primates) western gorilla	174
NP_008197.1	(primates) chimpanzee	174
NP_006939.1	(carnivores) harbor seal	175
NP_007080.1	(carnivores) gray seal	175
AP_004903.1	(rodents) Norway rat	172
NP_904339.1	(rodents) house mouse	172
NP_007405.1	(marsupials) common wallaroo	167



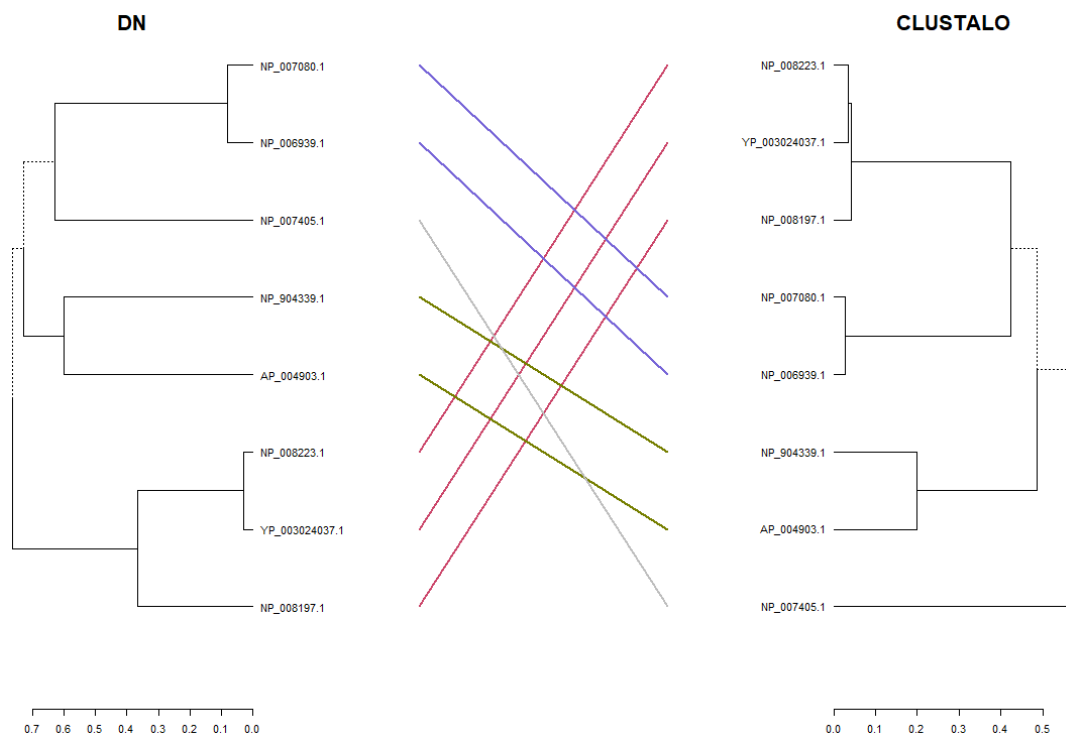
Dodatna Figura 4: Dendrogramski prikaz sekvenci proteina NADH dehidrogenaze 6 (MT-ND6) konstruisano R-P/F metodom za DN tip ponovaka. U listovima su prikazani identifikatori sekvenci, iza koje sledi informacija o vrsti kojoj ta sekvenca pripada



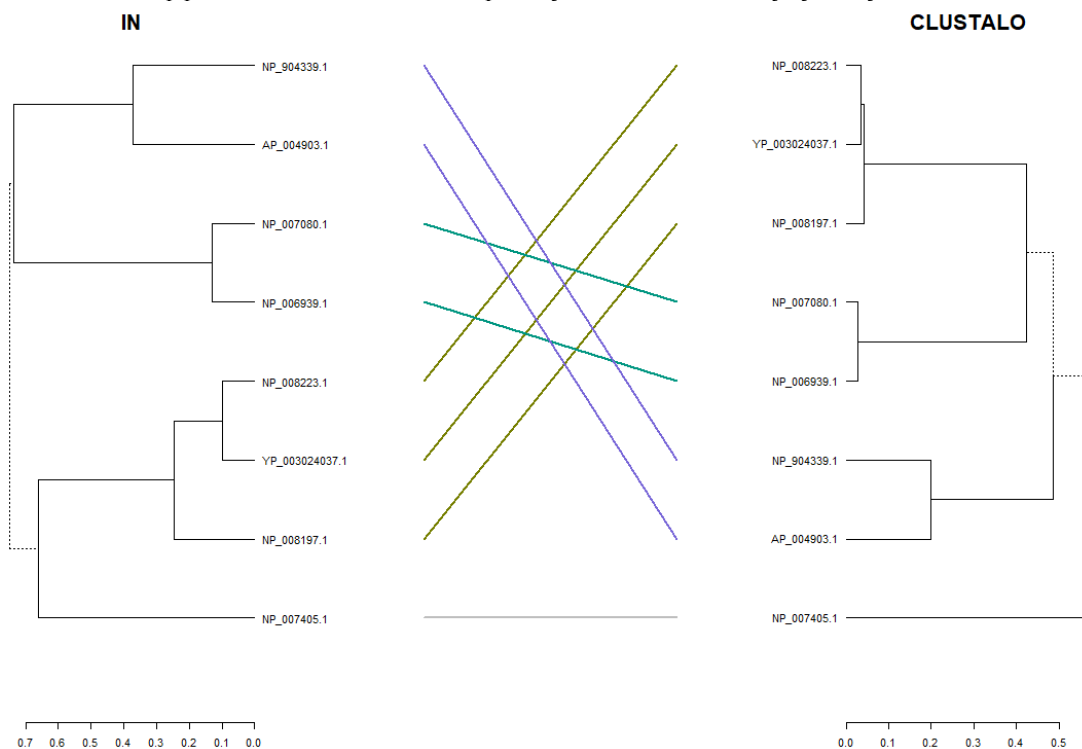
Dodatna Figura 5: Dendrogramski prikaz sekvenci proteina NADH dehidrogenaze 6 (MT-ND6) konstruisano R-P/F metodom za IN tip ponovaka. U listovima su prikazani identifikatori sekvenci, iza koje sledi informacija o vrsti kojoj ta sekvenca pripada



Dodatna Figura 6: Dendrogramski prikaz sekvenci proteina NADH dehidrogenaze 6 (MT-ND6) konstruisano UPGMA metodom primenjenom na matricu rastojanja dobijenu iz ClustalO.



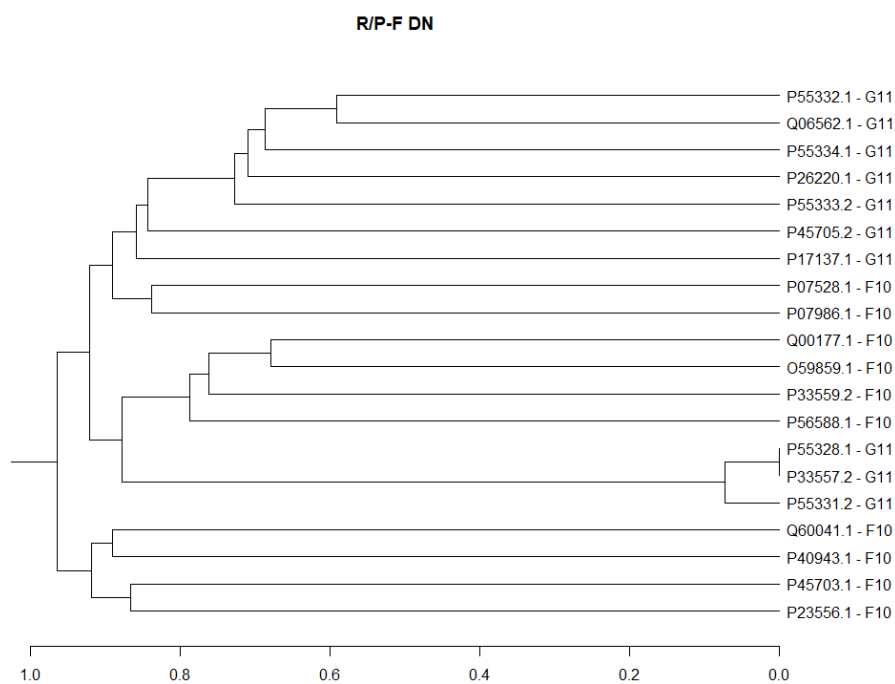
Dodatna Figura 7: Komparativni prikaz dendrograma sekvenci proteina NADH dehidrogenaze 6 (MT-ND6) konstruisano R-P/F metodom za DN tip ponovaka i UPGMA metodom primenjenom na matricu rastojanja dobijenu iz ClustalΩ.



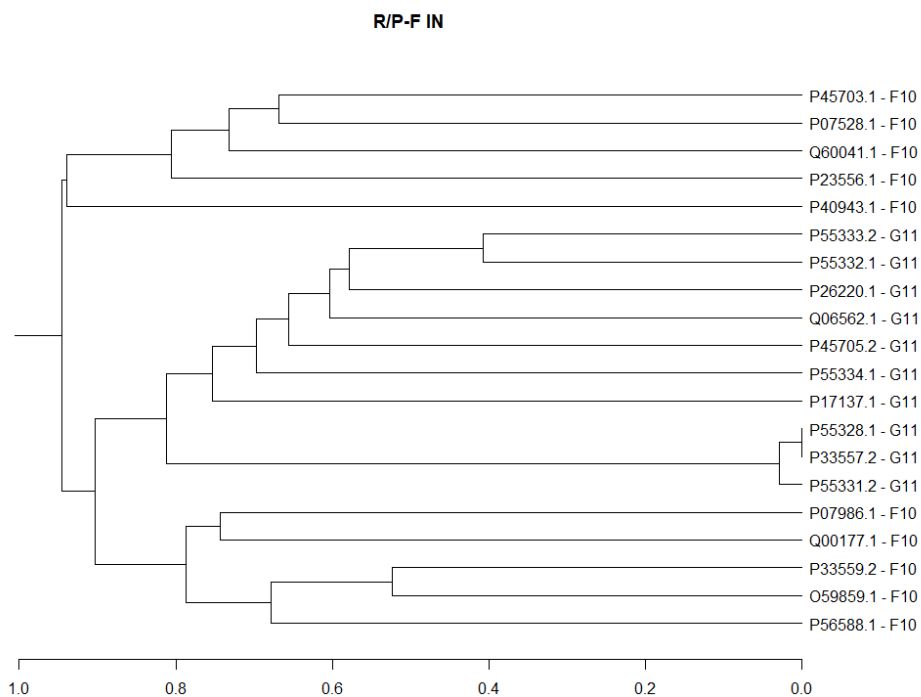
Dodatna Figura 8: Komparativni prikaz dendrograma sekvenci proteina NADH dehidrogenaze 6 (MT-ND6) konstruisano R-P/F metodom za IN tip ponovaka i UPGMA metodom primenjenom na matricu rastojanja dobijenu iz ClustalΩ.

Dodatna tabela 3: Sažete informacije o 20 proteinskih sekvenci skupa ksilanaze

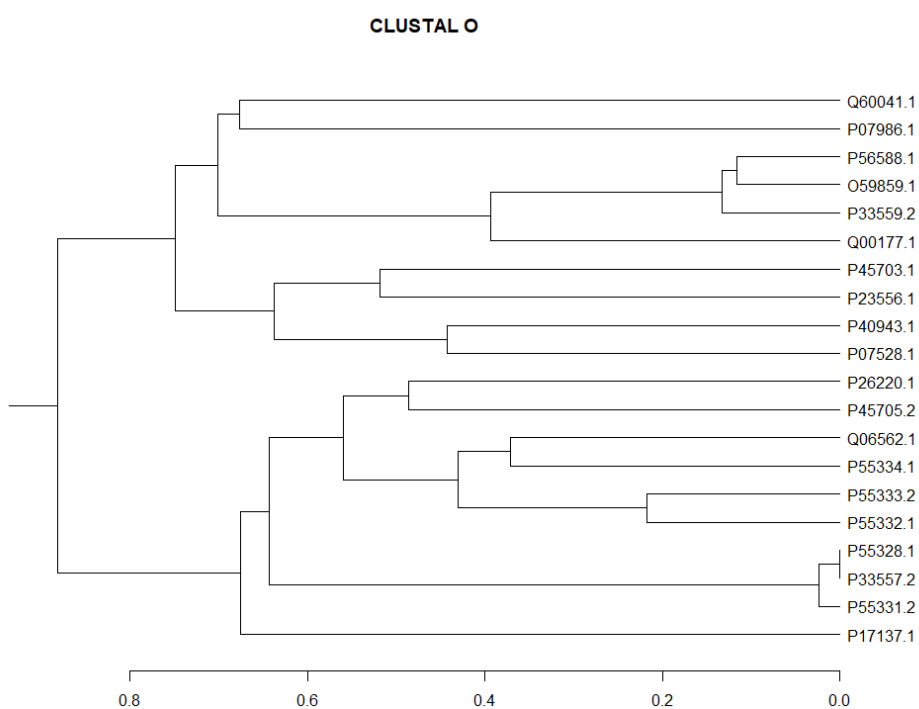
ID	Grupa	Dužina proteinske sekvence	ID	Grupa	Dužina proteinske sekvence
O59859.1	F10	327	P33557.2	G11	211
P56588.1	F10	302	P55328.1	G11	211
P33559.2	F10	327	P55331.2	G11	211
Q00177.1	F10	327	P45705.2	G11	210
P07986.1	F10	484	P26220.1	G11	240
P07528.1	F10	396	P55334.1	G11	227
P40943.1	F10	407	Q06562.1	G11	221
P23556.1	F10	342	P55332.1	G11	225
P45703.1	F10	330	P55333.2	G11	221
Q60041.1	F10	346	P17137.1	G11	261



Dodatna Figura 9: Dendrogramski prikaz sekvenci proteina ksilenaze konstruisano R-P/F metodom za DN tip ponovaka. U listovima su prikazani identifikatori sekvenci, iza koje sledi informacija o grupi kojoj ta sekvenca pripada

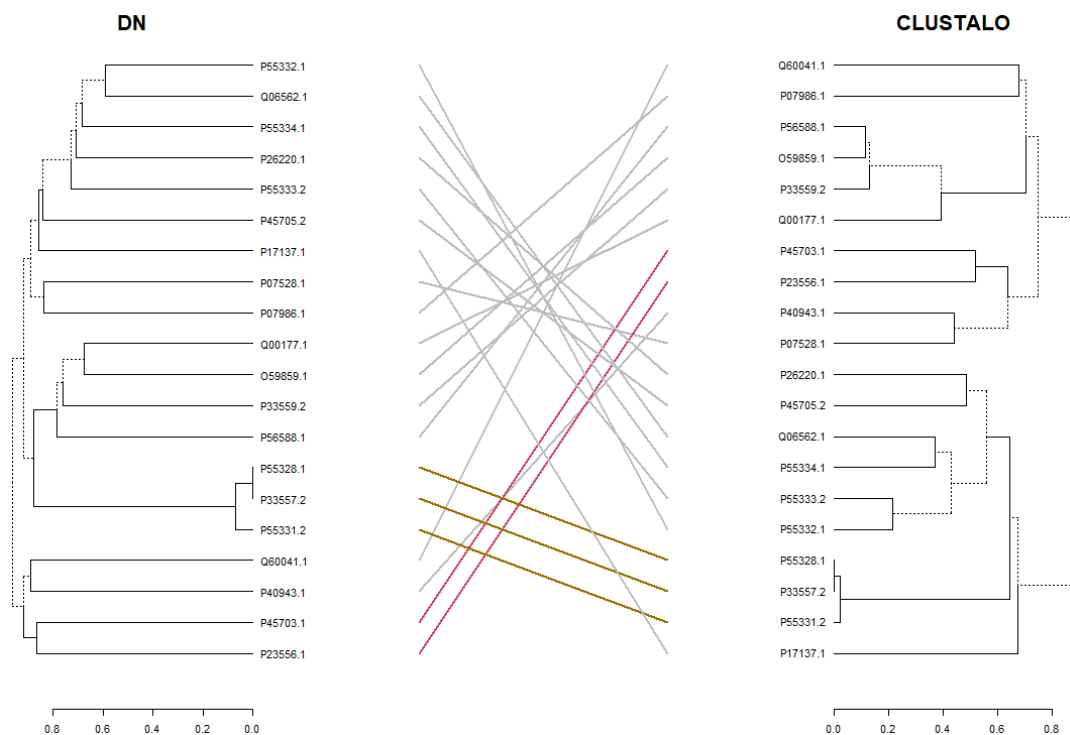


Dodatna Figura 10: Dendrogramski prikaz sekvenci proteina ksilenaze konstruisano R-P/F metodom za IN tip ponovaka. U listovima su prikazani identifikatori sekvenci, iza koje sledi informacija o grupi kojoj ta sekvenca pripada

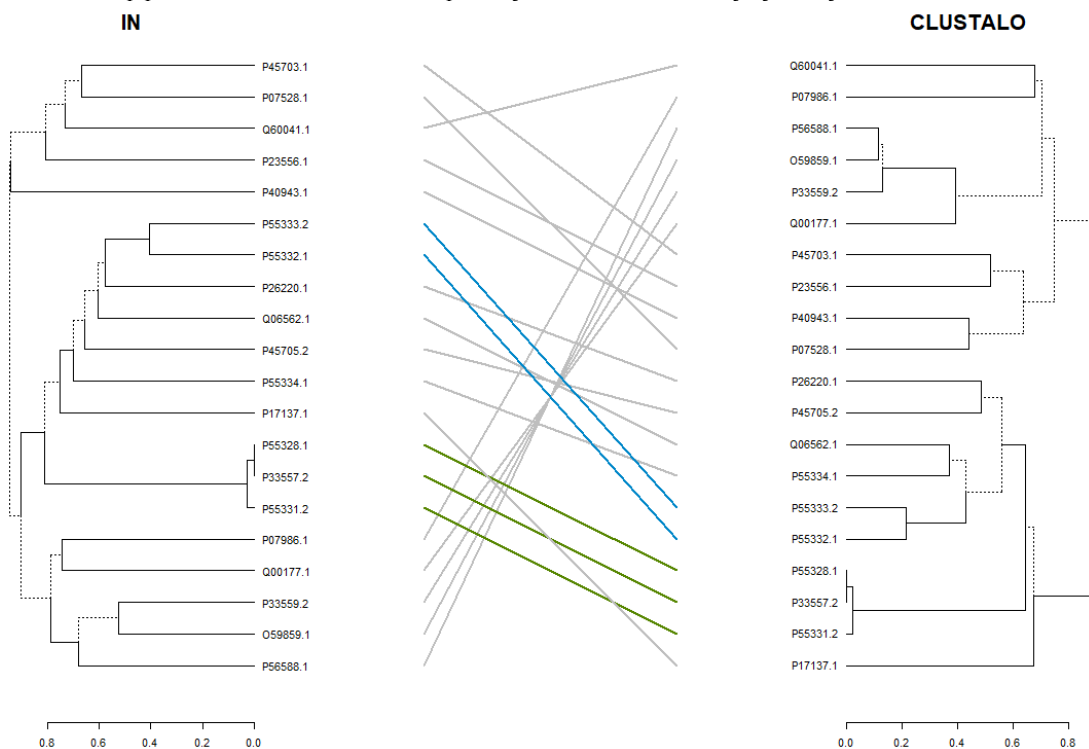


Dodatna Figura 11: Dendrogramski prikaz sekvenci proteina ksilenaze konstruisano UPGMA metodom primenjenom na matricu rastojanja dobijenu iz ClustalO.





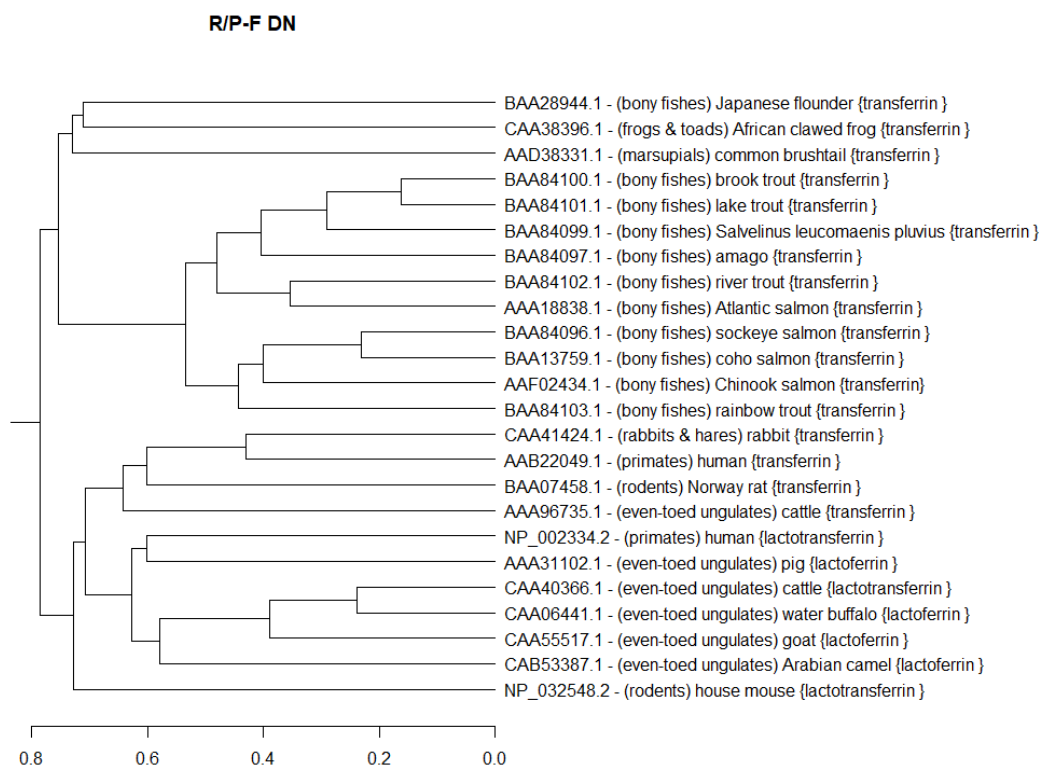
Dodatna Figura 12: Komparativni prikaz dendrograma sekvenci proteina ksilenaze konstruisano R-P/F metodom za DN tip ponovaka i UPGMA metodom primenjenom na matricu rastojanja dobijenu iz ClustalΩ.



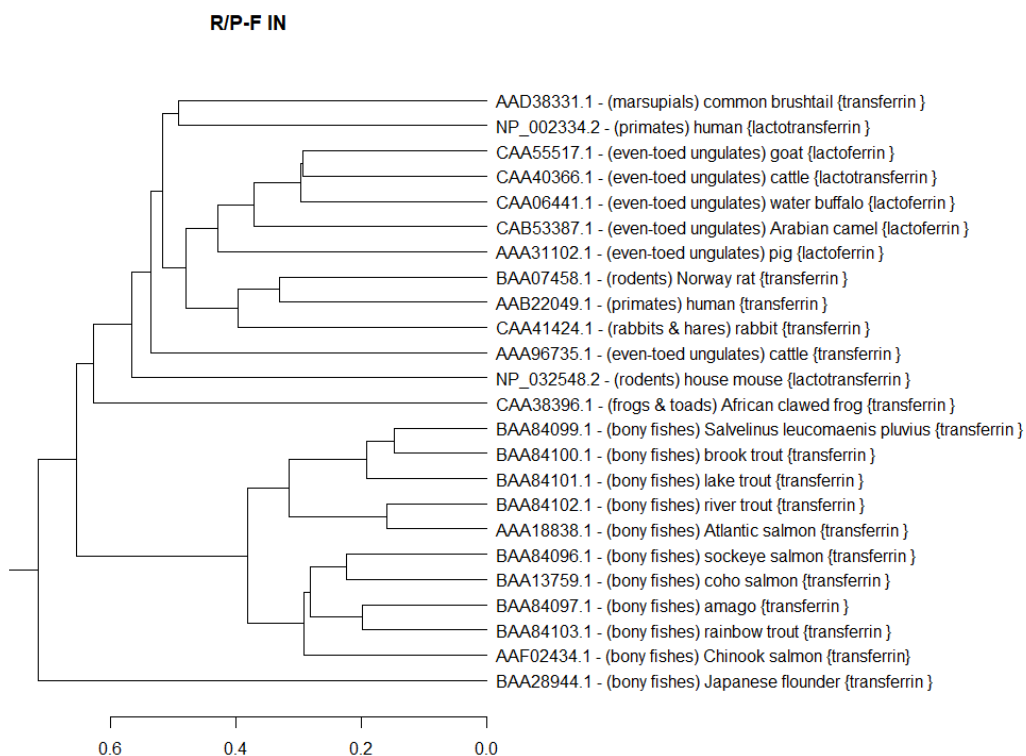
Dodatna Figura 13: Komparativni prikaz dendrograma sekvenci proteina ksilenaze konstruisano R-P/F metodom za IN tip ponovaka i UPGMA metodom primenjenom na matricu rastojanja dobijenu iz ClustalΩ.

Dodatna tabela 4: Sažete informacije o 24 proteinske sekvenci skupa proteina transferina

<b>ID</b>	<b>Protein</b>	<b>Red</b>	<b>Vrsta</b>	<b>Dužina proteinske sekvence</b>
AAB22049.1	transferrin	primates	human	698
CAA41424.1	transferrin	rabbits & hares	rabbit	694
BAA07458.1	transferrin	rodents	Norway rat	698
AAA96735.1	transferrin	even-toed ungulates	cattle	704
CAA06441.1	lactoferrin	even-toed ungulates	water buffalo	708
CAA40366.1	lactotransferrin	even-toed ungulates	cattle	708
CAA55517.1	lactoferrin	even-toed ungulates	goat	708
CAB53387.1	lactoferrin	even-toed ungulates	Arabian camel	708
AAA31102.1	lactoferrin	even-toed ungulates	pig	686
NP_002334.2	lactotransferrin	primates	human	710
NP_032548.2	lactotransferrin	rodents	house mouse	707
AAD38331.1	transferrin	marsupials	common brushtail	711
CAA38396.1	transferrin	frogs & toads	African clawed frog	717
BAA28944.1	transferrin	bony fishes	Japanese flounder	685
AAA18838.1	transferrin	bony fishes	Atlantic salmon	690
BAA84102.1	transferrin	bony fishes	river trout	691
BAA84101.1	transferrin	bony fishes	lake trout	691
BAA84100.1	transferrin	bony fishes	brook trout	691
BAA84099.1	transferrin	bony fishes	Salvelinus leucomaenis pluvius	691
AAF02434.1	transferrin	bony fishes	Chinook salmon	677
BAA13759.1	transferrin	bony fishes	coho salmon	687
BAA84096.1	transferrin	bony fishes	sockeye salmon	691
BAA84103.1	transferrin	bony fishes	rainbow trout	691
BAA84097.1	transferrin	bony fishes	amago	691

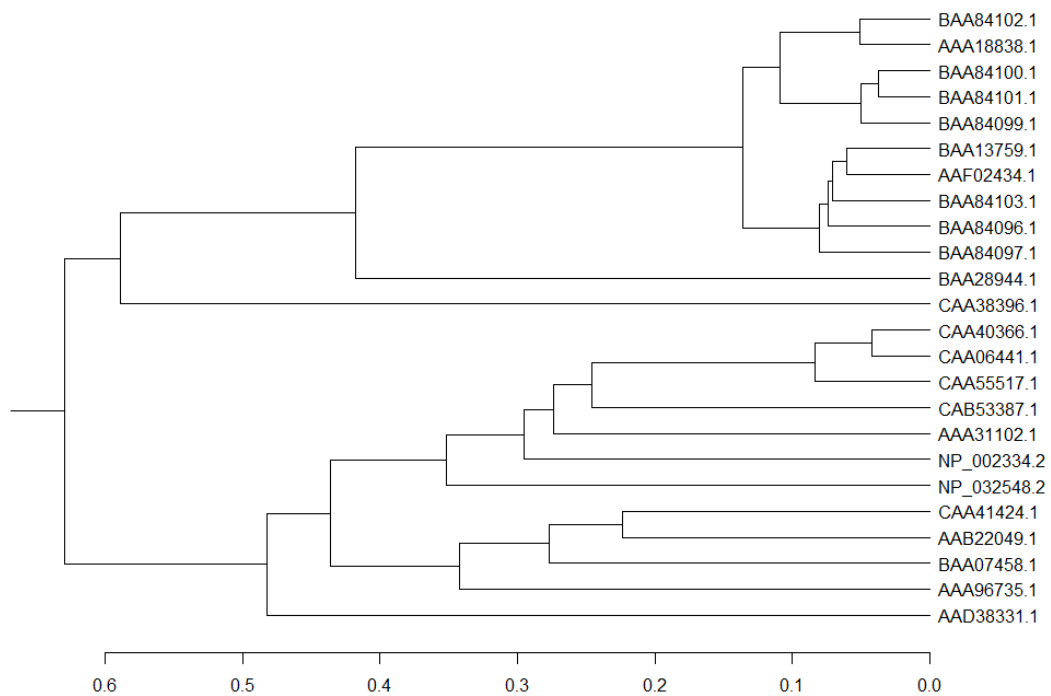


Dodatna Figura 14: Dendrogramski prikaz sekvenci proteina transferina konstruisano R-P/F metodom za DN tip ponovaka. U listovima su prikazani identifikatori sekvenci, iza koje slede dodatne informacije o vrsti proteinima kojoj te sekvence pripadaju

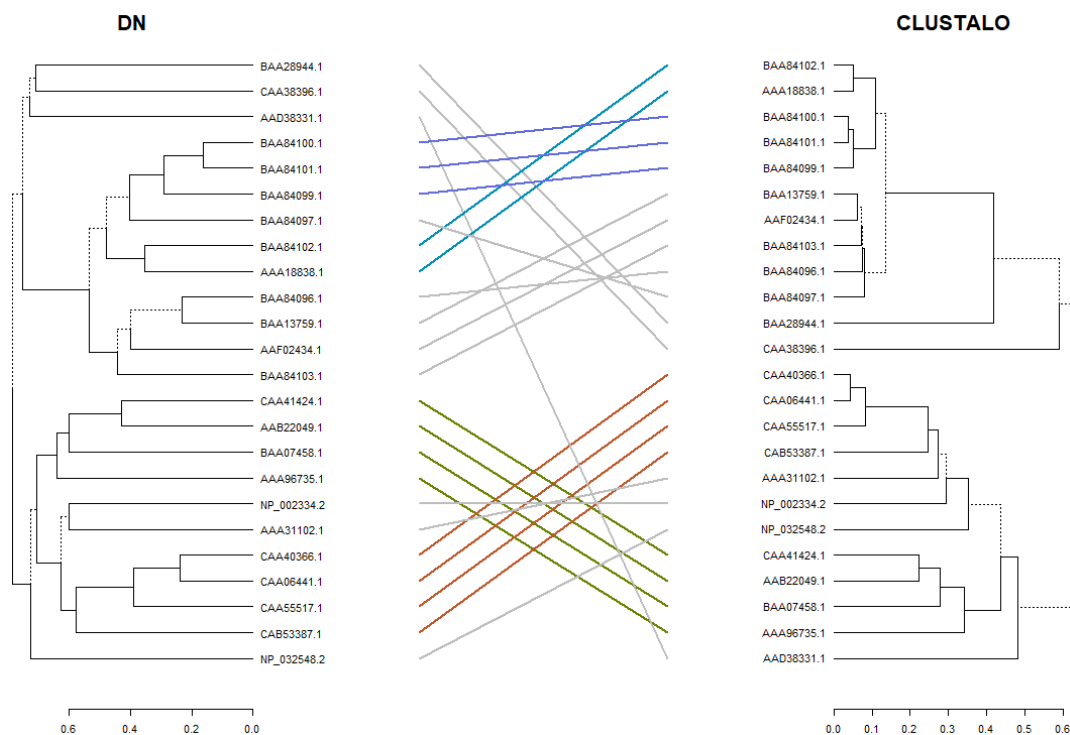


Dodatna Figura 15: Dendrogramski prikaz sekvenci proteina transferina konstruisano R-P/F metodom za IN tip ponovaka. U listovima su prikazani identifikatori sekvenci, iza koje slede dodatne informacije o vrsti proteinima kojoj te sekvence pripadaju

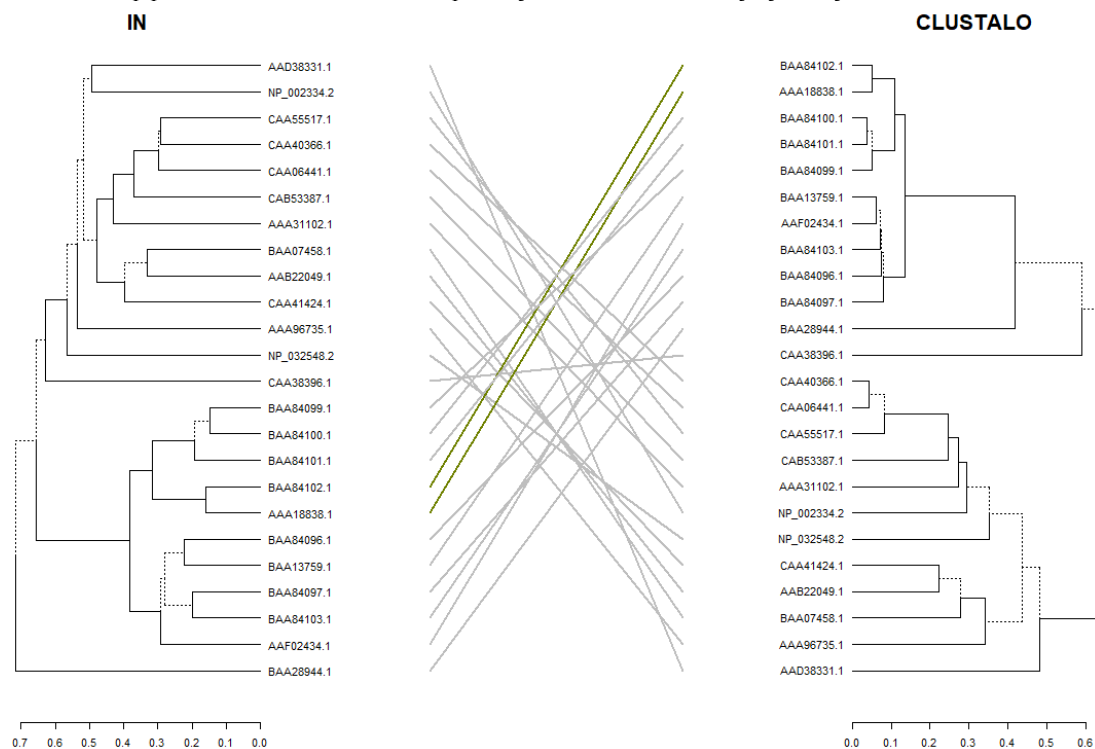
# CLUSTAL O



Dodatna Figura 16: Dendrogramski prikaz sekvenci proteina transferina konstruisano UPGMA metodom primjenenom na matricu rastojanja dobijenu iz ClustalO.



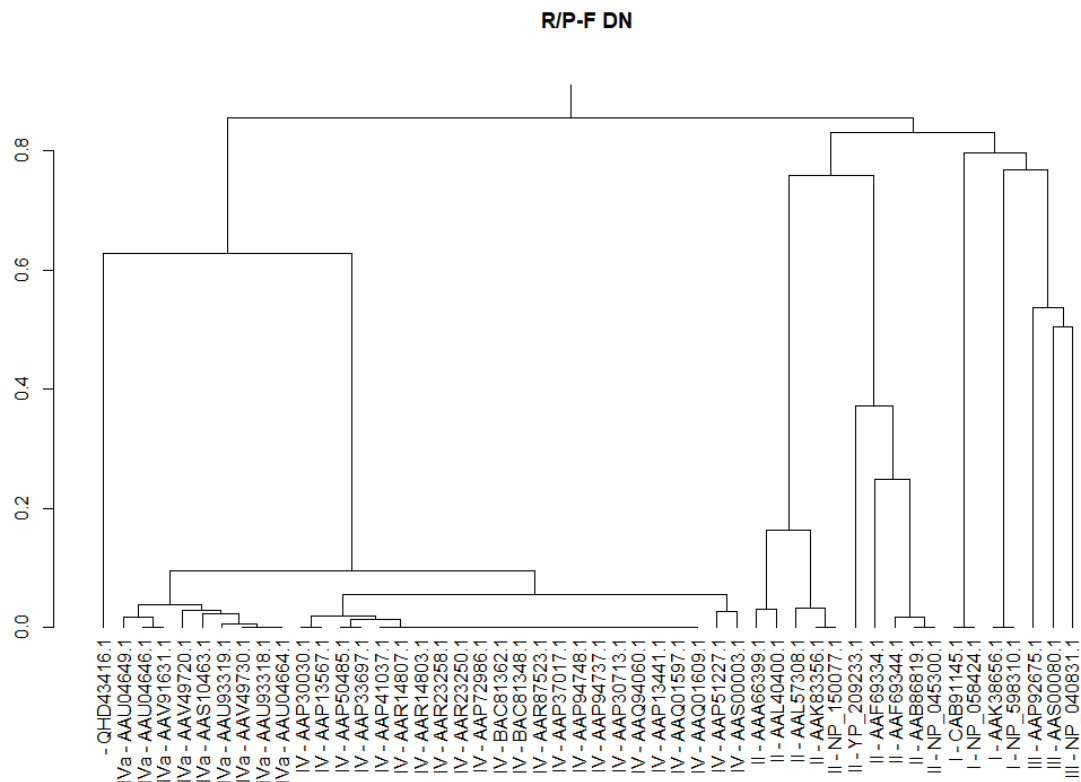
Dodatna Figura 17: Komparativni prikaz dendrograma sekvenci proteina transferina konstruisano R-P/F metodom za DN tip ponovaka i UPGMA metodom primenjenom na matricu rastojanja dobijenu iz ClustalΩ.



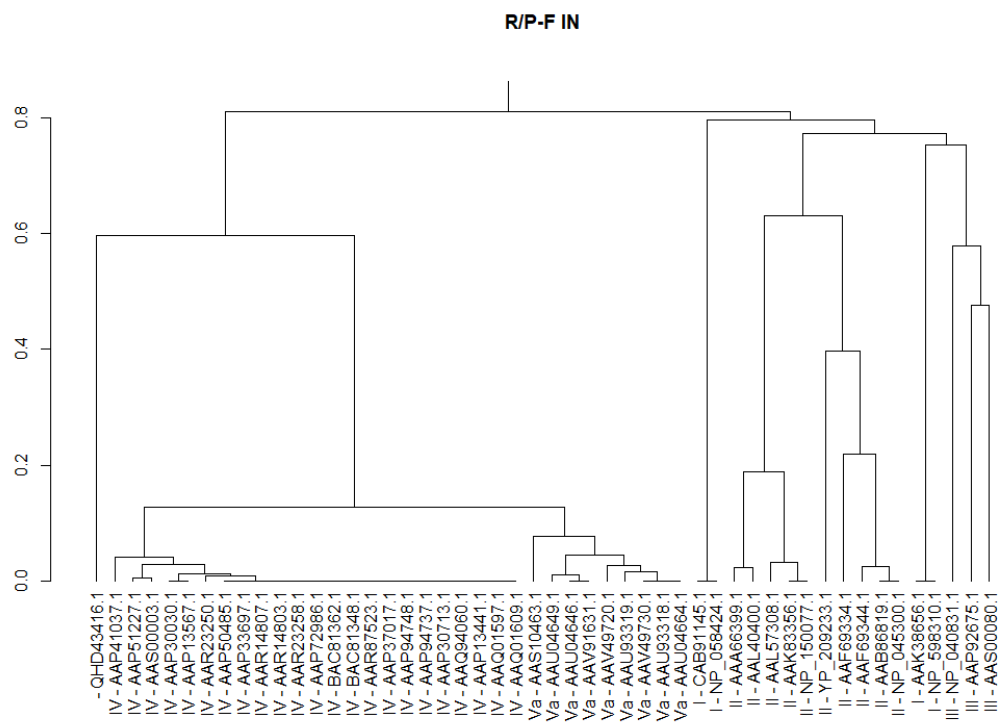
Dodatna Figura 18: Komparativni prikaz dendrograma sekvenci proteina transferina konstruisano R-P/F metodom za IN tip ponovaka i UPGMA metodom primenjenom na matricu rastojanja dobijenu iz ClustalΩ.

Dodatna tabela 5: Sažete informacije o 50 proteinske sekvenci skupa spajk proteina korona virusa

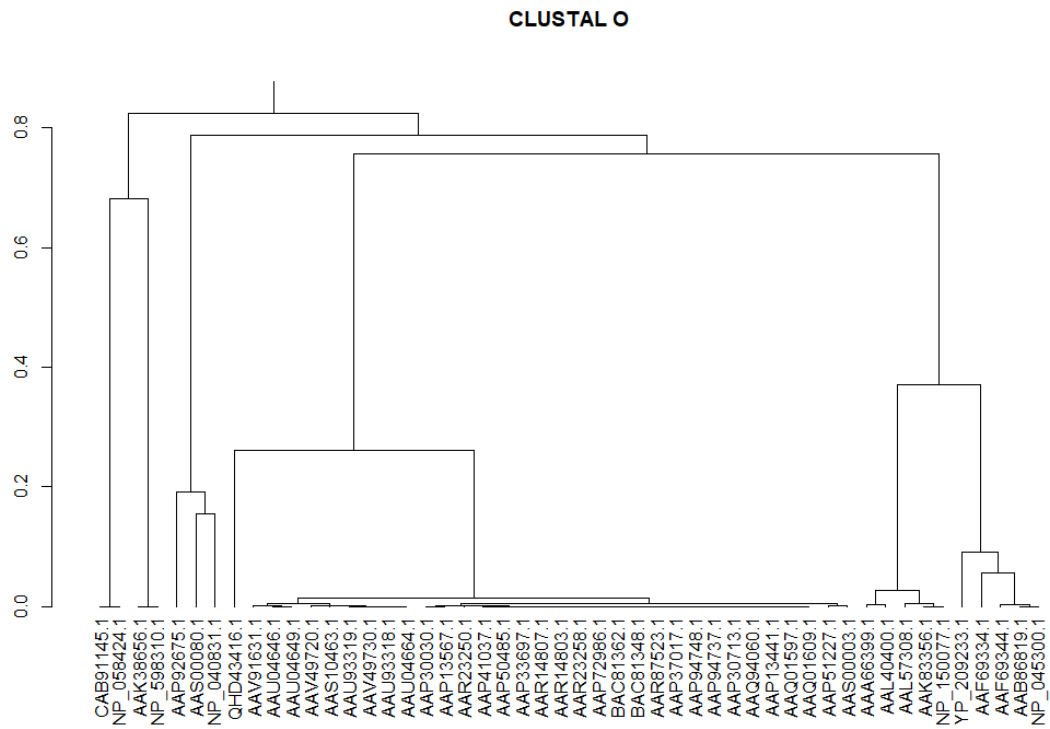
ID	Grupa	Dužina proteinske sekvence	ID	Grupa	Dužina proteinske sekvence
CAB91145.1	I	1447	AAV49730.1	IVa	1255
NP_058424.1	I	1447	AAP51227.1	IV	1255
AAK38656.1	I	1383	AAS00003.1	IV	1255
NP_598310.1	I	1383	AAP30030.1	IV	1255
AAK83356.1	II	1363	AAP13567.1	IV	1255
AAL57308.1	II	1363	AAP50485.1	IV	1255
AAA66399.1	II	1363	AAP41037.1	IV	1255
AAL40400.1	II	1363	AAQ01597.1	IV	1255
NP_150077.1	II	1363	AAQ01609.1	IV	1255
AAB86819.1	II	1324	AAP13441.1	IV	1255
YP_209233.1	II	1376	AAQ94060.1	IV	1255
AAF69334.1	II	1321	AAP30713.1	IV	1255
AAF69344.1	II	1324	AAP33697.1	IV	1255
NP_045300.1	II	1324	AAP94737.1	IV	1255
AAP92675.1	III	1169	AAP94748.1	IV	1255
AAS00080.1	III	1169	AAP37017.1	IV	1255
NP_040831.1	III	1162	AAR87523.1	IV	1255
AAS10463.1	IVa	1255	BAC81348.1	IV	1255
AAU93318.1	IVa	1255	BAC81362.1	IV	1255
AAV49720.1	IVa	1255	AAP72986.1	IV	1255
AAU93319.1	IVa	1255	AAR23250.1	IV	1255
AAU04646.1	IVa	1255	AAR23258.1	IV	1255
AAU04649.1	IVa	1255	AAR14803.1	IV	1255
AAU04664.1	IVa	1255	AAR14807.1	IV	1255
AAV91631.1	IVa	1255	QHD43416.1	*	1273
grupe I i II sadrže koronaviruse sisara, grupa III sadrži korona viruse ptica i grupa IV sadrži SARS-CoVs viruse * spike protein Covid-19 SARS-CoV-2 virusa					



Dodatna Figura 19: Dendrogramski prikaz sekvenci spajk proteina korona virusa konstruisano R-P/F metodom za DN tip ponovaka. U listovima su prikazani identifikatori sekvenci, iza koje sledi informacija o grupi kojoj ta sekvenca pripada



Dodatna Figura 20: Dendrogramski prikaz sekvenci spajk proteina korona virusa konstruisano R-P/F metodom za IN tip ponovaka. U listovima su prikazani identifikatori sekvenci, iza koje sledi informacija o grupi kojoj ta sekvenca pripada



Dodatna Figura 21: Dendrogramski prikaz sekvenci spajk proteina korona virusa konstruisano UPGMA metodom primenjenom na matricu rastojanja dobijenu iz ClustalΩ.

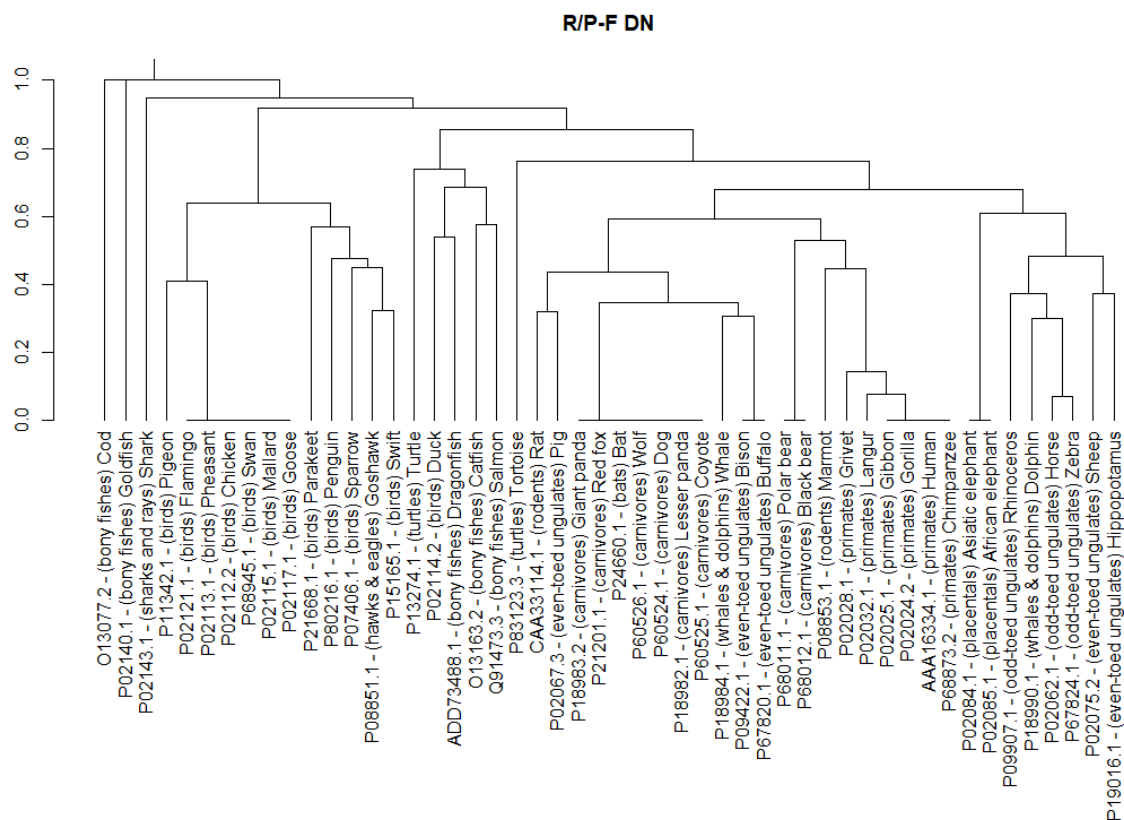




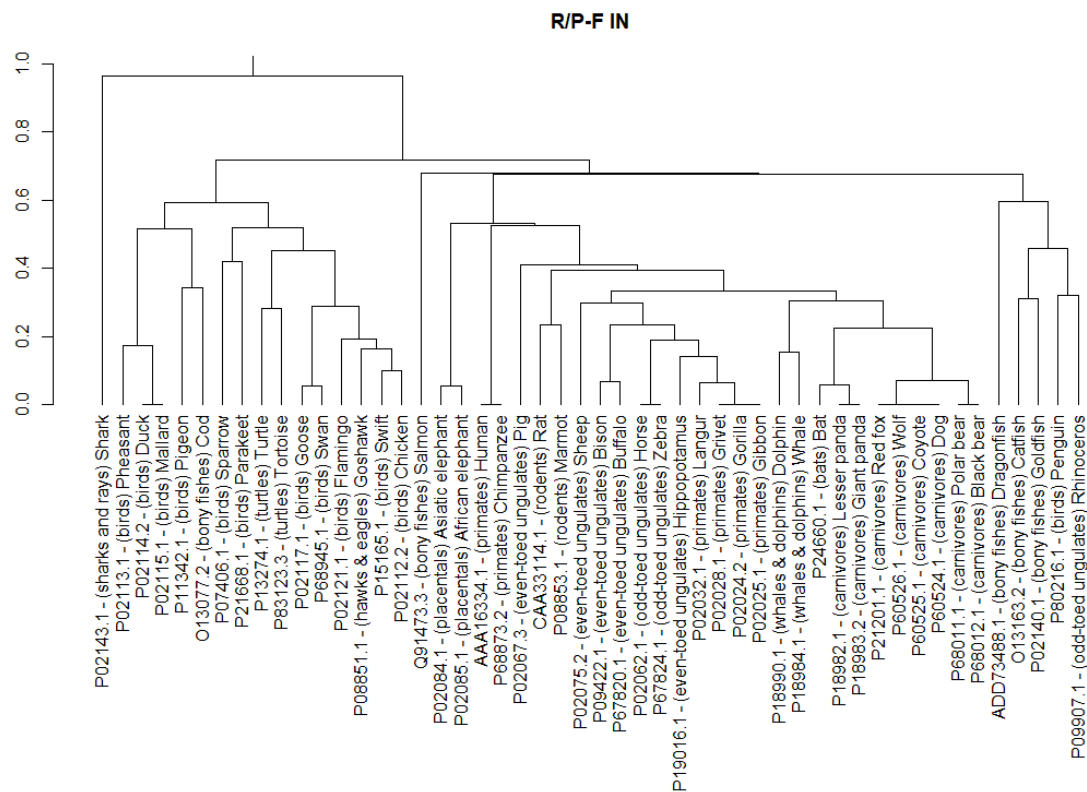
Dodatna tabela 5: Sažete informacije o 50 proteinske sekvenci proteina beta globina

ID	Vrsta	Dužina proteinske sekvence
AAA16334.1	(primates) Human	147
P08851.1	(hawks & eagles) Goshawk	146
P18982.1	(carnivores) Lesser panda	146
P18983.2	(carnivores) Giant panda	147
P02075.2	(even-toed ungulates) Sheep	145
P02114.2	(birds) Duck	147
P02115.1	(birds) Mallard	146
P02117.1	(birds) Goose	146
CAA33114.1	(rodents) Rat	147
P80216.1	(birds) Penguin	146
P15165.1	(birds) Swift	146
P60525.1	(carnivores) Coyote	146
O13163.2	(bony fishes) Catfish	148
P09422.1	(even-toed ungulates) Bison	145
P68945.1	(birds) Swan	146
P67820.1	(even-toed ungulates) Buffalo	145
P60524.1	(carnivores) Dog	146
P68873.2	(primates) Chimpanzee	147
P18990.1	(whales & dolphins) Dolphin	146
P02140.1	(bony fishes) Goldfish	147
P68011.1	(carnivores) Polar bear	146
P09907.1	(odd-toed ungulates) Rhinoceros	146
P02112.2	(birds) Chicken	147
P60526.1	(carnivores) Wolf	146
P13274.1	(turtles) Turtle	146
P11342.1	(birds) Pigeon	146
P68012.1	(carnivores) Black bear	146
P02084.1	(placentals) Asiatic elephant	146
P02085.1	(placentals) African elephant	146
P83123.3	(turtles) Tortoise	147
P02028.1	(primates) Grivet	146
P02024.2	(primates) Gorilla	147
P02143.1	(sharks and rays) Shark	141
P19016.1	(even-toed ungulates) Hippopotamus	146
P02062.1	(odd-toed ungulates) Horse	146
P02025.1	(primates) Gibbon	146
P18984.1	(whales & dolphins) Whale	146
P24660.1	(bats) Bat	146
P21201.1	(carnivores) Red fox	146
P08853.1	(rodents) Marmot	146
Q91473.3	(bony fishes) Salmon	148

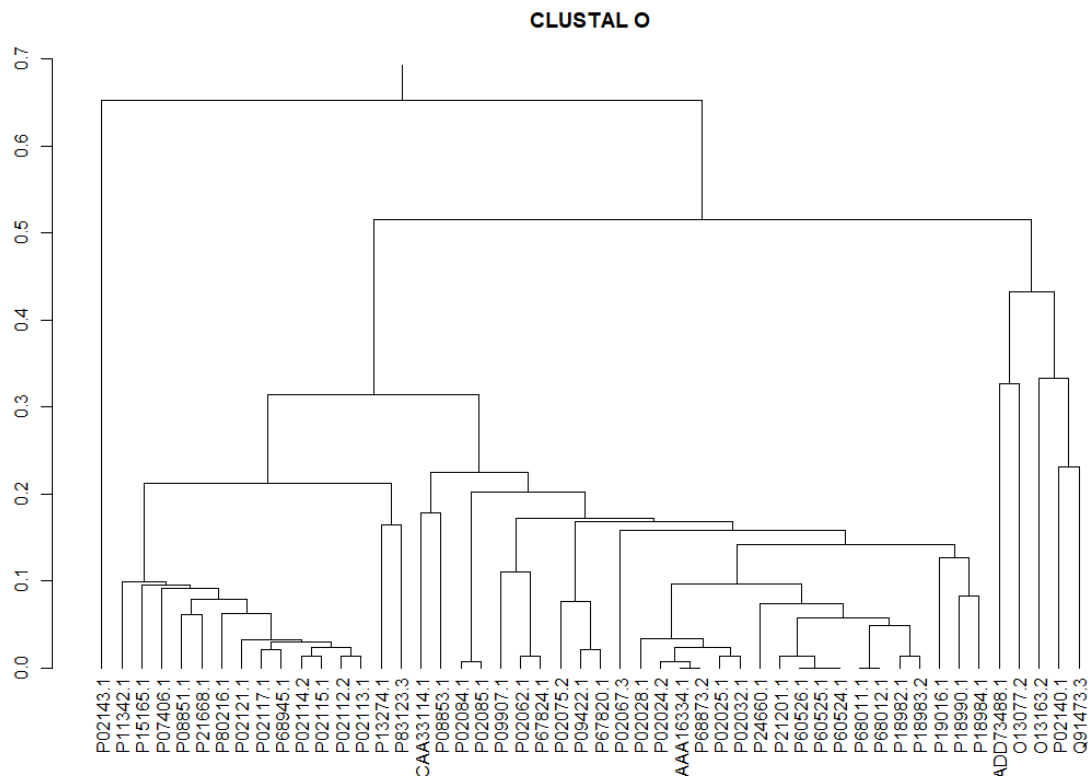
P07406.1	(birds) Sparrow	146
P02113.1	(birds) Pheasant	146
P02121.1	(birds) Flamingo	146
P02067.3	(even-toed ungulates) Pig	147
ADD73488.1	(bony fishes) Dragonfish	141
P21668.1	(birds) Parakeet	146
P67824.1	(odd-toed ungulates) Zebra	146
O13077.2	(bony fishes) Cod	147
P02032.1	(primates) Langur	146



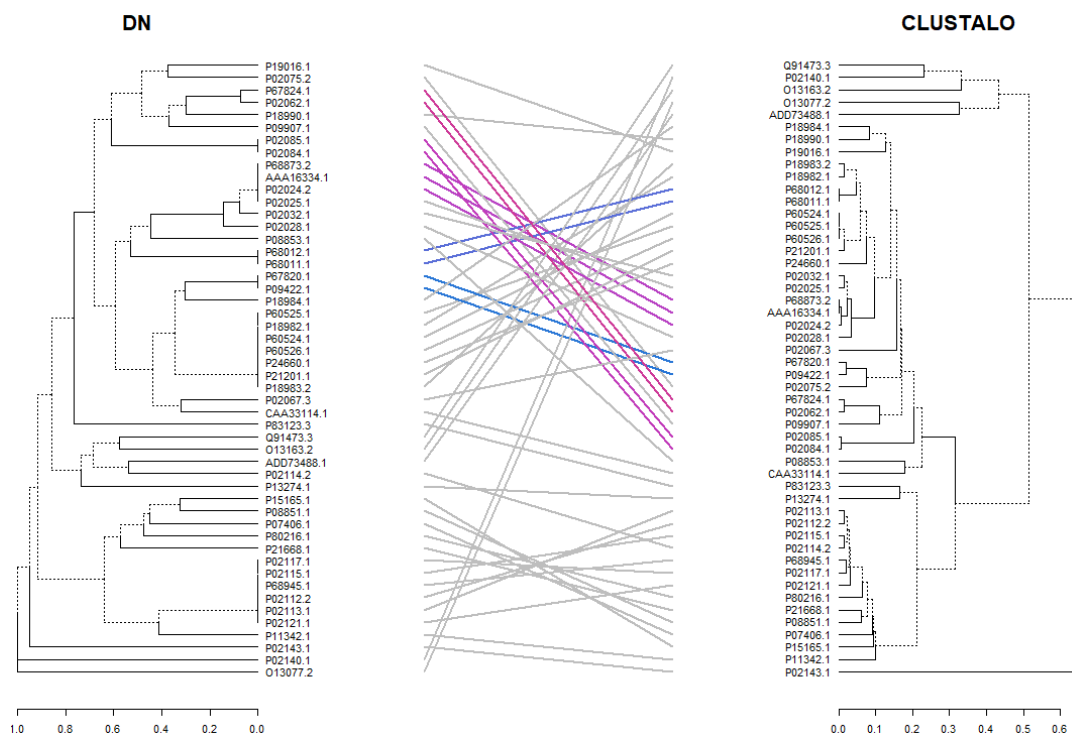
Dodatna Figura 24: Dendrogramski prikaz sekvenci beta globina konstruisano R-P/F metodom za DN tip ponovaka. U listovima su prikazani identifikatori sekvenci, iza koje sledi informacija vrsti kojoj ta sekvenca pripada



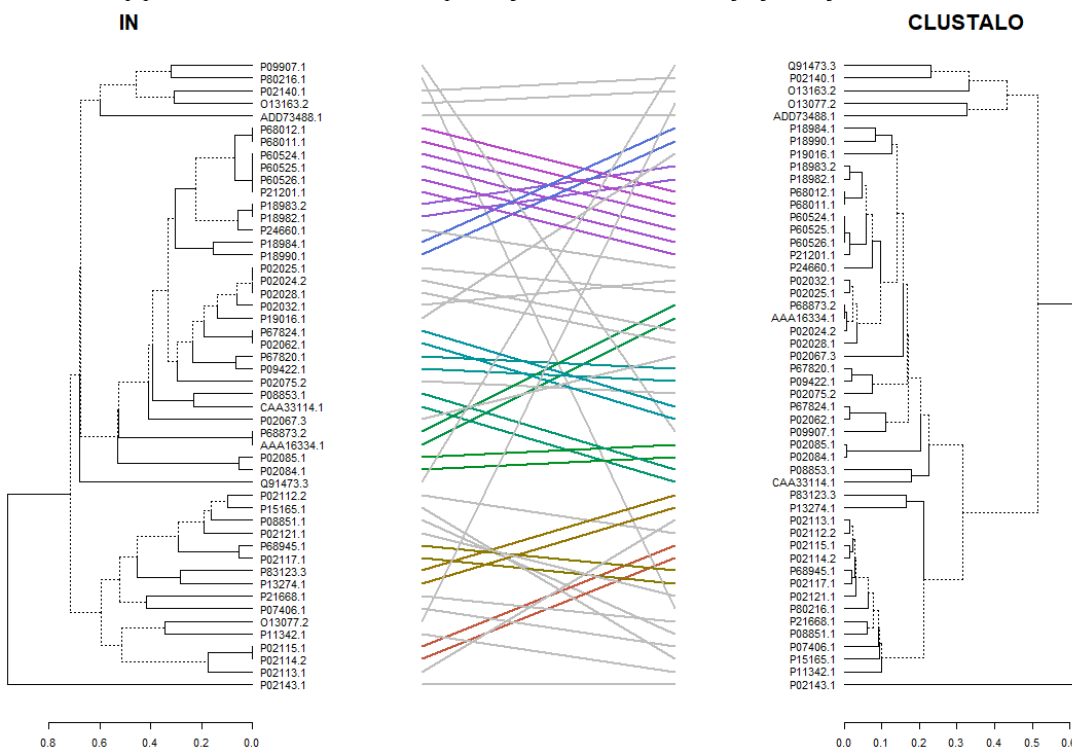
Dodatna Figura 25: Dendrogramski prikaz sekvenci proteina beta globina konstruisano R-P/F metodom za IN tip ponovaka. U listovima su prikazani identifikatori sekvenci, iza koje sledi informacija o vrsti kojoj ta sekvenca pripada



Dodatna Figura 26: Dendrogramski prikaz sekvenci proteina beta globina konstruisano UPGMA metodom primenjenom na matricu rastojanja dobijenu iz ClustalQ.



Dodatna Figura 27: Komparativni prikaz dendrograma sekvenci proteina beta globina konstruisano konstruisano R-P/F metodom za DN tip ponovaka i UPGMA metodom primenjenom na matricu rastojanja dobijenu iz ClustalΩ.



Dodatna Figura 28: Komparativni prikaz dendrograma sekvenci proteina beta-globina konstruisano konstruisano R-P/F metodom za IN tip ponovaka i UPGMA metodom primenjenom na matricu rastojanja dobijenu iz ClustalΩ.