Phylogenetic study on the evolution of opine dehydrogenases

Sonia Chothani
Carnegie Mellon University

12-5-2012

Abstract

Opines are chemical compounds found in crown-gall and hairy tumors. The biosynthesis genes are inserted into the plant genome by Agrobacteria while infecting plants. Although this is the most well-known source of opines, they were first found in Octopus muscle and various such contexts unrelated to plants' crown-gall or tumor. A very interesting phylogenetic question intrigued me as to how did these opines evolve, whether there was an early on duplication event or is it an example of convergent evolution. A systematic phylogenetic analysis was carried out. The sequence alignment revealed structural motifs possibly related to the active sites of the enzyme. WAG + I + G model was used to generate trees for this alignment. The tree analysis compared and annotated sequences in six taxonomic classes, including 51 species, representing different Opine dehydrogenases. The tree clustered the enzymes according to the evolutionary relationships of the species. Thus this study helped understand the phylogeny of opine dehydrogenases. At the same time a striking horizontal gene transfer from a bacteria, Lactobacillus ruminis to the marine algae, Nannochloropsis gaditana was also spotted out.

I. Introduction

"Opines are low molecular weight compounds found in crown gall and hairy tumors"[1]. The agrobacteria tricks the plants to produce these for its nitrogen and energy demands. These agrobacteria have opine synthesis genes which they insert into the plant host genome while infecting[1]. Although this is the most widely known source of opines, surprisingly they are present and were first discovered in octopus muscle (known as *Octopine*). They are also present in various other marine invertebrates (known as *Alanopine, Tauropine*). They can also be formed in normal callus and plant tissue by arginine metabolism (as *acetopine*, *nopaline*). Saccharopine occurs in fungi, and higher mammals too[2]!

The obvious phylogenetic question one would raise here as to where did this variety arise from? The probable hypotheses could be either there were gene duplications early on and then followed by speciation events leading to this presence of opines in variety of species or it could be that there was a speciation event first and then in each of those sub-trees there were independent duplication events arising to this biochemical variety of opines and hence we could be looking at another wonderful example of convergent evolution. It could in fact be a hybrid of these two hypotheses. To further explore this question I carried out a systematic phylogenetic study using the procedure as explained in the methods section.

One of the most important and well studied opine biosynthesis gene is the opine dehydrogenase. This enzyme acts on the CH-NH substrate bond using NAD(+) or NADP(+) as an acceptor[3]. As explained earlier these are also present in bacteria and marine cephalods. Bacteria transfer a portion of this plasmid (T-DNA) to a susceptible plant cell while infecting, the T-DNA then integrates into the plant nuclear genome, where its genes can be expressed.[3].

Specifically these genes direct the synthesis of opines in plants which the bacteria further uses as its nitrogen source. In marine invertebrates, its activity in the muscle is significantly correlated with a species' ability to buffer the acidic end products of anaerobic metabolism[4]. Here, they mainly catalyze the terminal step of anaerobic glycolysis during muscle anoxia associated with locomotion[5].

II. Methods

Selecting opine dehydrogenase from all the biosynthesis genes was a choice based on a few factors. Firstly, since the dataset involved species from very diverged groups it would be difficult to analyze the whole set of biosynthesis genes, hence focusing on one of the biosynthesis genes would be an ideal start. Secondly, sequence data of dehydrogenases was available for numerous species. Thus the opine dehydrogenases were selected for this analysis. Further the Amino Acid sequences were used, as they would give us more information in this case considering the diversity of species.

The sequences were acquired by a BLAST carried out at threshold 10 on the Opine Dehydrogenase sequence (Haliotis discus hannai) acquired from UniProt *(Refer Appendix A1)*. The sequences were further screened based on the Taxonomy report, hence selecting atleast one representative from each taxa and while doing so the factors like number of hits, e-value were taken into consideration. GOLD was also used to check the quality of some of the sequences in the list. In this way the bias and redundancy was reduced as much as possible.

A final data set of 52 sequences was generated. The sequences were renamed using a Perl script to get the 4 letter protein name combined with 5 letter for specie name. Hence making it easy to read and similar to the Uniprot standard. These were further visualized in Jalview and a screening of redundant sequences by setting a cutoff of 99% was carried out.

Further, Multiple sequence alignments were carried out through Jalview using MAFFT, MUSCLE, ProbCons and T-Coffee. This alignments were analyzed for which program worked best. The criteria used was the active site information available from UniProt for some of the opine dehydrogenases. MAFFT had done a relatively better job and hence I further trimmed this MAFFT alignment manually and introduced gaps to improve the alignment in relation to the active sites. Further all the columns with >50% gaps were trimmed. This trimmed alignment was used for future analysis.*(Refer Appendix A2)*

This alignment was loaded in TOPALI to find the best model for this alignment. Both by AIC1 and BIC the WAG+I+G model was best. Hence I chose to go ahead with that model.
The following were the parameters of the best model generated by TOPALI's model generator. *For more details refer Appendix A3*

Table 1: *Model parameters*

WAG + I + G		
AIC1,BIC	alpha	pINV
47587.36, 48000.78	3.397	0.009

These parameters were used to generate a Maximum Likelihood tree using the WAG model in Seaview. All other parameters were kept as default. I also generated a Neighbor Joining tree with 1000 replicates using Seaview to see its performance on this alignment. The trees were saved in newick unrooted format and also the squared and circular format. *(Refer Appendix A4 & A5)*

Next step was to generate a specie list, i.e a list of all the species involved in this alignment *(Refer Appendix A6)*. This list was uploaded on NCBI taxonomy common tree to obtain a specie tree. This tree was viewed using Notung-2.6 and color annotated according to Bacteria,

Gastropoda, Octopodia, Decapodiforms, Pectinidae, Polychaeia. These annotations were saved as a text file. It is important to note that NCBI is conservative while generating these trees and hence the tree is non-binary, which means where the branching is uncertain, it is left as a polytomy. *(Refer Appendix A7)*

Further the gene tree obtained from seaview was viewed in Notung-2.6 and the color coded specie annotations were imported into this tree. Since this gene tree obtained by Maximum likelihood method is unrooted, it was rooted under the rooting tab in Notung by selecting the bacteria subtree. This is because we would expect this subtree to branch out first. After which this was reconciled with the specie tree. This step helped obtain possible duplication/losses, hence in understanding the phylogeny of events that could have occurred. *(Refer Appendix A8)*Although an intriguing grouping of a bacteria and a marine algae motivated to repeat this step with the "infer transfers" tab option in Notung-2.7-Beta. Hence the reconcilation was carried out again with transfers included at a cost of 4.0. *(Refer Appendix A9)*

Hence a good amount of data to answer the question was generated and the results and analysis are further discussed in the next section.

III. Results and Discussion

A multiple sequence alignment of the 52 sequences obtained after screening was analyzed for the possible binding site motifs. The two sites on UniProt given for Octopine Dehydrogenase in Pectin Maximus was highly conserved in the sequence alignment. Specifically Arg-324 and His-212. Further reading up the literature on NMR studies on Pectin Maximus' Octopine dehydrogenase, reconfirmed that Arginine-324 plays an important role in the enzyme's function[6]. This along with other such observed conserved motifs could be an insight to an interesting enzyme mechanism conserved over diverged species.

Further the model generator indicated WAG + I + G model to be the best by both AIC1 and BIC parameters. These parameters as shown in methods section were used to generate maximum likelihood trees in seaview. When compared with neighbor joining tree these trees seemed to have performed better. This is because it clearly discriminated the various taxa present whereas NJ tree had distances much closer even for bacteria and other species in picture. Hence the ML tree was used for further analysis. Very clearly the bootstrap values also well supported the branches, giving more confidence on the tree.

In Notung-2.6 the gene tree was reconciled with the specie tree which gave interesting results to answer our question. Most of the species were well represented in accordance with their biological evolutionary relationship as expected from the specie tree. Hence hinting on a convergent evolution example. The branches related to both these contexts have strong bootstrap values. Although it seemed to have overestimated the duplications in bacteria and this could be simply due to weak signal/error. Nevertheless since the question here was to understand relation of bacteria with respect to metazoa this was not a problem. There was an early on duplication observed which means somewhere this gene was introduced and over the course of evolution the other species lost the gene, the losses can also be observed in the notung tree. Further the tree with transfers inferred seemed to give a more clear picture of this relationship. Here it seemed more like an independent duplication and transfers which lead to the introduction of the gene. These results now allow for a further detailed study on other opine biosynthesis genes, which would give more insight on the evolution of opines as a whole in diverged species. Overall this study gave a good start to understanding the evolution of opines in

diverged species.

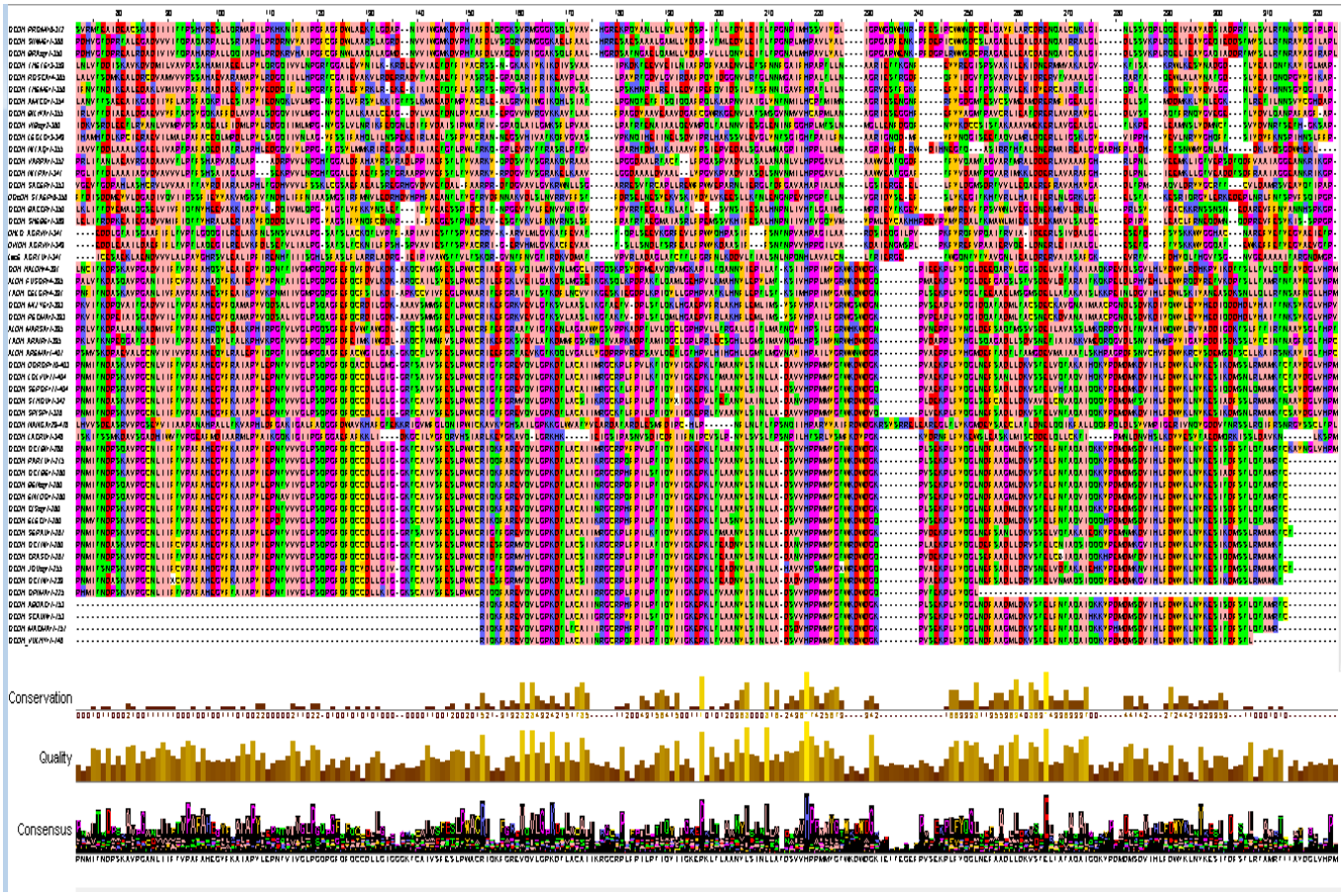
Another direction this study lead to is the transfer observed from a bacteria (Lactobacillus ruminis) to a marine algae (Nannochloropsis gaditana) which was very striking. Literature on the algae makes it seem quite a plausible situation as it shows Nannochloropsis has exhibited involvement in various other horizontal gene transfers[7].[8].

V. APPENDIX

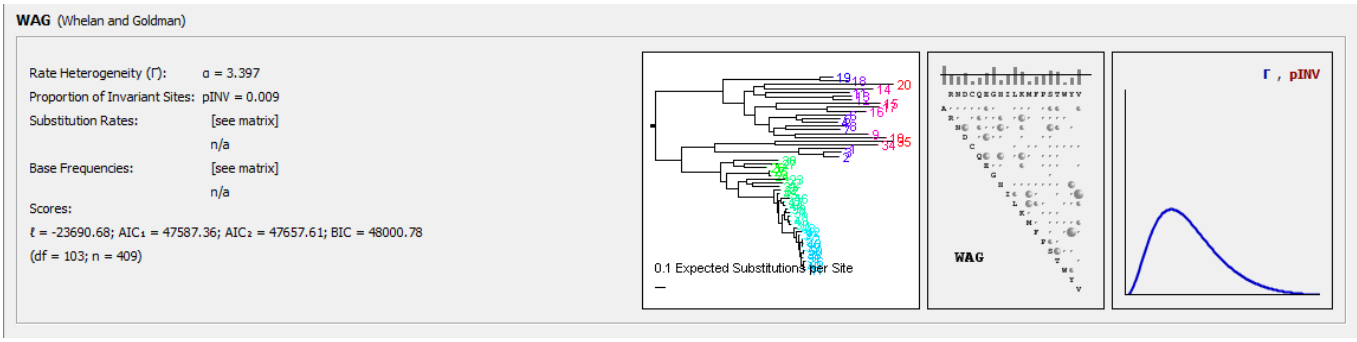
A1: Fasta Sequence of the sequence used in BLAST to obtain more sequences for the dataset

```
>sp|Q8N0N9|ODH_HALDH Opine dehydrogenase OS=Haliotis discus hannai GN=tadh
PE=2 SV=1
MTKKITVLVCGGGNGAHVTAGLAASRDDIETRVLTTFADEAERWTNIMKENDLRITVDEG
DIKSGESVDFKVKLNCITKDPSKAVPGADVIIFTVPAFAHQSYLEAIEPYIQPNTTIVGM
PGQPGFEFQVFDVLKDKAKQCVIMSFESLPWACRIAEFGKFVQILMVKVNLMGCLIRGQS
KPSYDPMEAVQRVMGKAPILTQANNYIEPIIATKSIIHPPIMYGKWKDWDGKPIEEKPLF
YQGLDEEQARYLGGISDELVATAKAIAAQKPEVDLSGVLHLYDWYLRDHKPYIKDTTSLL
TVLQTDTAYDGLVHPMKETEDGKFVPDFRYRYLTEDVPNGLVVTRGLAQIAGVPTPYHDE
VIAWCQKQLGKEIIVGDELKGKDIGSTRCPQRYGINTMDALVNIM
```

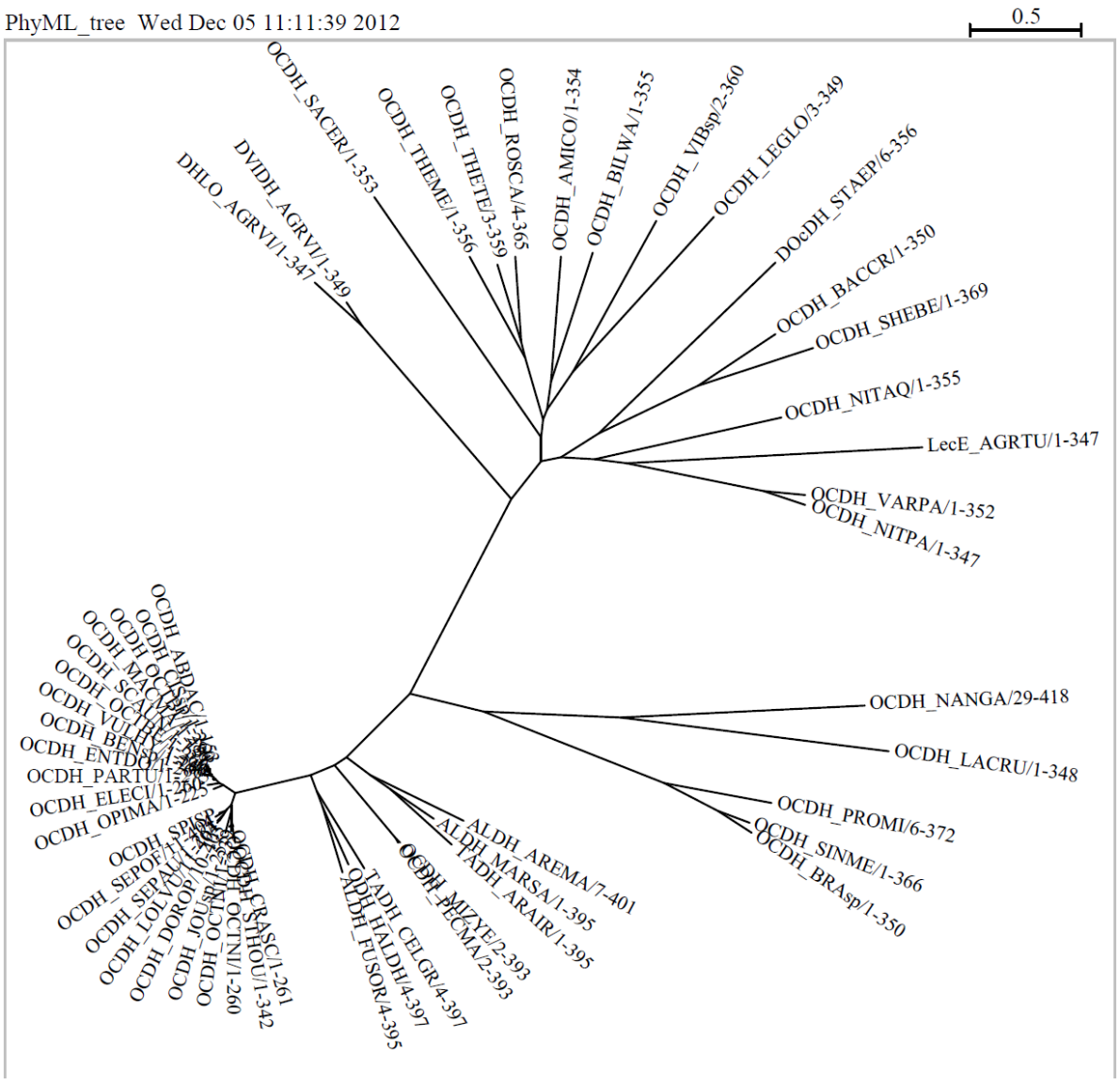
A2: Trimmed MAFFT Multiple sequence alignment



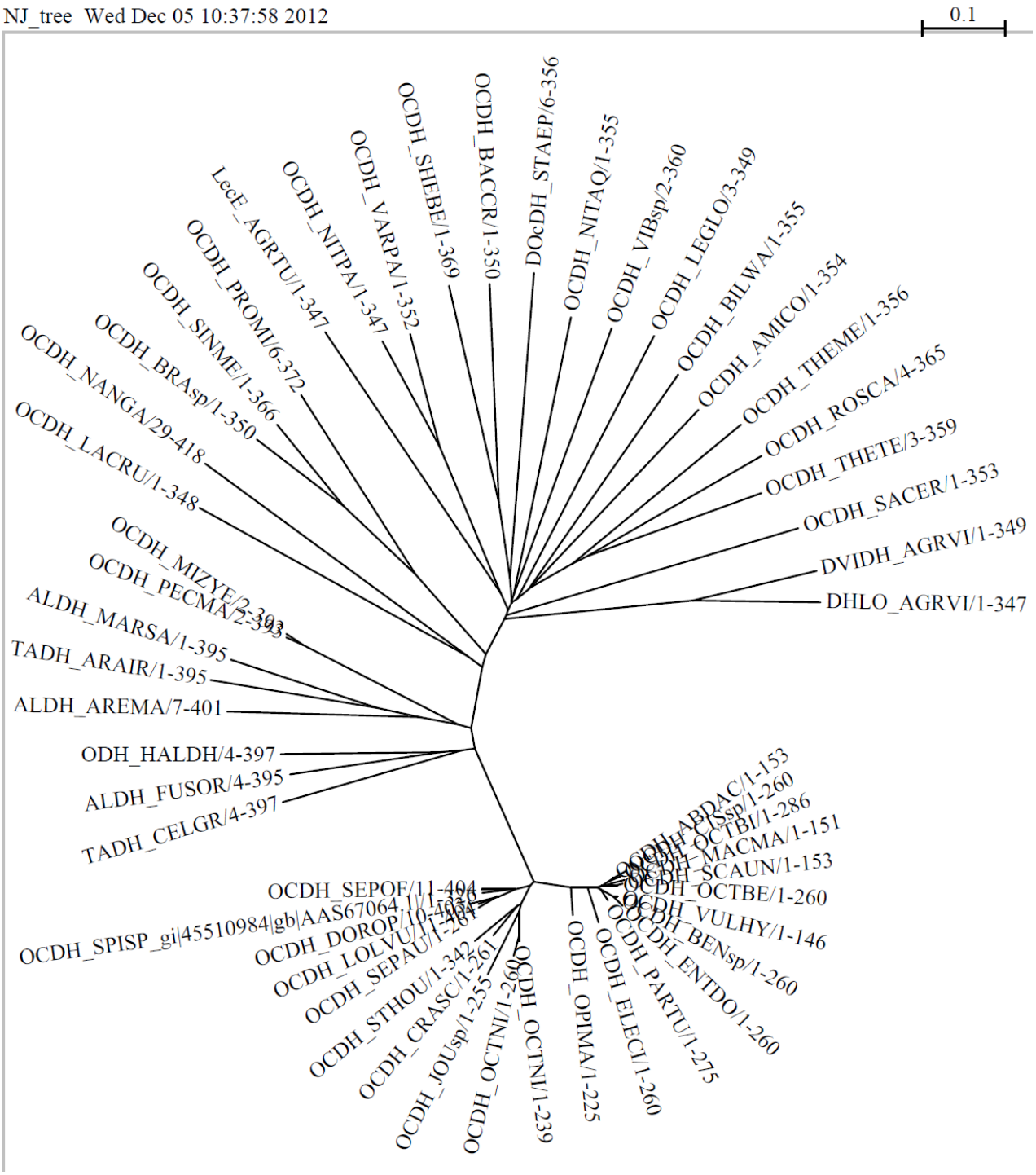
A3: Model Generator – Best Model



A4: Maximum Likelihood tree generated using WAG+I+G Model



A5: Neighbor Joining tree generated using 1000 replicates



A6: Specie List

Proteus mirabilis
Bradyrhizobium
Roseiflexus castenholzii
Aminobacterium colombiense
Vibrio
Nitratireductor aquibiodomus
Nitratireductor pacificus
Staphylococcus epidermidis
Shewanella benthica
Agrobacterium tumefaciens
Fusitriton oregonensis
Mizuhopecten yessoensis
Marphysa sanguinea
Arenicola marina
Loligo vulgaris
Sthenoteuthis oualaniensis
Nannochloropsis gaditana
Octopus bimaculoides
Octopus berrima
Enteroctopus dofleini
Eledone cirrhosa
Octopoteuthis nielseni
Joubiniteuthis
Opisthoteuthis massyae
Scaeurgus unicirrhus
Vulcanoctopus hydrothermalis
Sinorhizobium meliloti
Thermoanaerobacter tengcongensis
Thermosipho melanesiensis
Bilophila wadsworthia
Legionella longbeachae
Variovorax paradoxus
Saccharopolyspora erythraea
Bacillus cereus
Agrobacterium vitis
Haliotis discus hannai
Cellana grata
Pecten Maximus
Arabella iricolor
Doryteuthis opalescens
Sepia officinalis
Spirula spirula
Lactobacillus ruminis
Pareledone turqueti
Benthoctopus sp.
Cistopus
Sepioteuthis australis
Cranchia scabra
Octopoteuthis nielseni
Abdopus aculeatus
Macroctopus maorum

A7: NCBI generated and Notung-2.6 color annotated specie tree

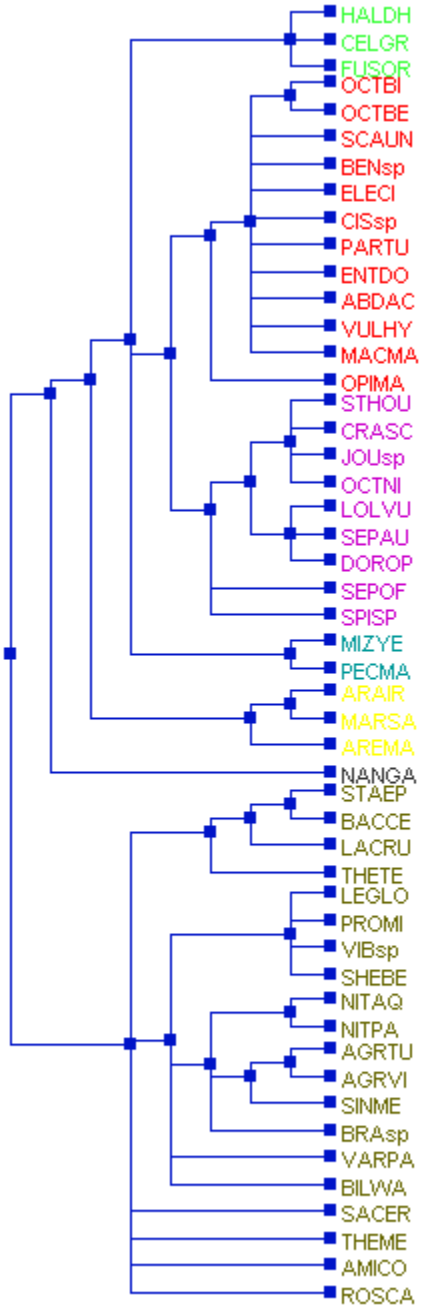


Image generated with Notung 2.7, on Dec 8, 2012

A8: Gene tree reconciled showing Duplications (Losses ignored for better visualization) with the A6 specie tree

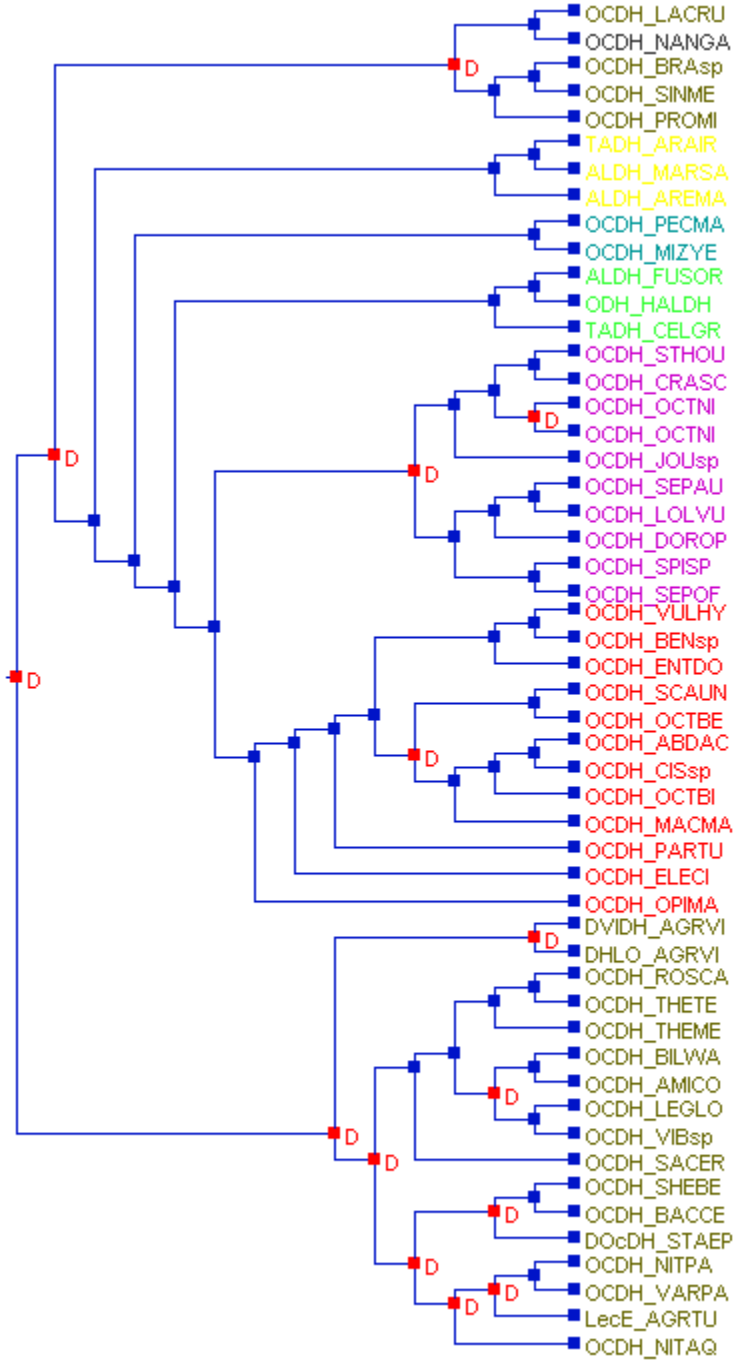


Image generated with Notung 2.7, on Dec 8, 2012

A9: Gene tree reconciled with transfers at a cost of 4.0 using Notung-2.7-beta with the A6 specie tree

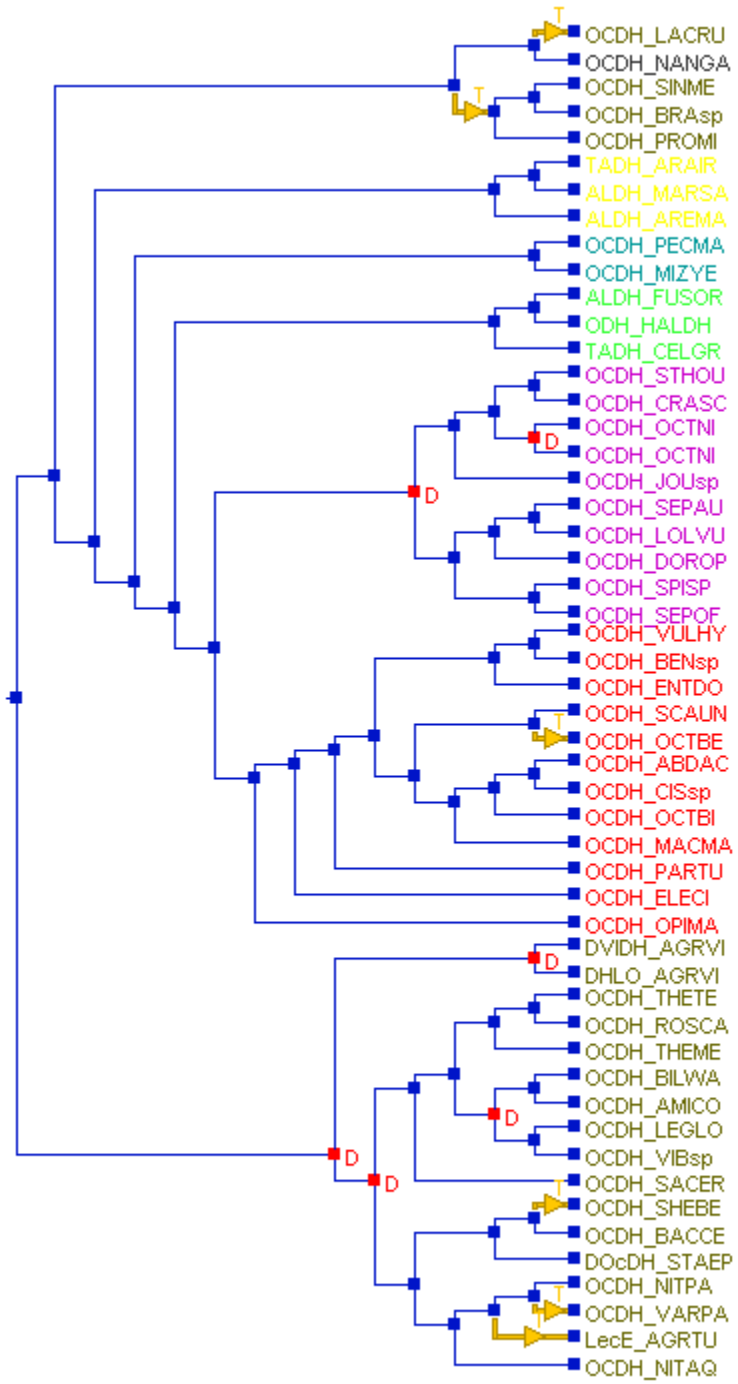


Image generated with Notung 2.7, on Dec 11, 2012

VI Acknowledgements

I would like to thank Dr. Dannie Durand for guiding me throughout the procedure and equipping me with these phylogenetic analyses tools. I feel confident to carry out a phylogenetic study post this course. I would also like to thank Pieter Spealman for various ideas and guidance during the office hours. Lastly the dataset I used for this project which was quite well-behaved!

VII REFERENCES

[1] *Kim, K. S., & Farrand, S. K. (1996)* Ti plasmid-encoded genes responsible for catabolism of the crown gall opine mannopine by Agrobacterium tumefaciens are homologs of the T-region genes responsible for synthesis of this opine by the plant tumor. *Journal of bacteriology, 178(11), 3275–84*

[2] *Moore, L. W., & Chilton, W. S. (1997)* Diversity of Opines and Opine-Catabolizing Bacteria Isolated from Naturally Occurring Crown Gall Tumors. *Appl. Environ. Microbol. 63(1):201*

[3] EMBL – EBI : Opine dehydrogenase (IPR003421)

[4] *Seibel, B. a, Thuesen, E. V., & Childress, J. J. (2000).* Light-limitation on predator-prey interactions: consequences for metabolism and locomotion of deep-sea cephalopods. *The Biological bulletin*, *198*(2), 284–98.

[5] *Balowin, J. (1983).* The Properties and Functions of Alanopine Dehydrogenase and Octopine Dehydrogenase from the Pedal Retractor Muscle of Strombidae. *Pacific Science 36(3).*

[6] *Smits, S. H. J., Meyer, T., Mueller, A., van Os, N., Stoldt, M., Willbold, D., Schmitt, L., et al. (2010).* Insights into the mechanism of ligand binding to octopine dehydrogenase from Pecten maximus by NMR and crystallography. *PloS one*, *5*(8), e12312. doi:10.1371/journal.pone.0012312

[7] *Minorsky, P. V. (2012).* On the Inside. *Plant Physiology, 158(4), 1485–1486. doi:10.1104/pp.112.900433*

[8] *Vieler, A., Wu, G., Tsai, C.-H., Bullard, B., Cornish, A. J., Harvey, C., Reca, I.-B., et al. (2012).* Genome, Functional Gene Annotation, and Nuclear Transformation of the Heterokont Oleaginous Alga Nannochloropsis oceanica CCMP1779. *(D. Bhattacharya, Ed.)PLoS Genetics, 8(11), e1003064. doi:10.1371/journal.pgen.1003064*