Pacific Symposium on Biocomputing 15:10-20(2010)

BENCHMARKING BLAST ACCURACY OF GENUS/PHYLA CLASSIFICATION OF METAGENOMIC READS

STEVEN D. ESSINGER AND GAIL L. ROSEN

Electrical & Computer Engineering, Drexel University, 3141 Chestnut Street
Philadelphia, PA 19141, USA

Metagenomics is the study of environmental samples. Because few tools exist for metagenomic analysis, a natural step has been to utilize the popular homology tool, BLAST, to search for sequence similarity between sample fragments and an administered database. Most biologists use this method today without knowing BLAST's accuracy, especially when a particular taxonomic class is under-represented in the database. The aim of this paper is to benchmark the performance of BLAST for taxonomic classification of metagenomic datasets in a supervised setting; meaning that the database contains microbes of the same class as the 'unknown' query fragments. We examine well- and under-represented genera and phyla in order to study their effect on the accuracy of BLAST. We conclude that on fine-resolution classes, such as genera, the accuracy of BLAST does not degrade very much with under-representation, but in a highly variant class, such as phyla, performance degrades significantly. Our analysis includes five-fold cross validation to substantiate our findings.

1. Introduction

The relatively new field of metagenomics has been rapidly expanding over the past several years [1, 2]. This field focuses on DNA obtained from an environmental sample rather than from pure cultures in a laboratory. This markedly substantial difference from conventional microbial genomics poses a unique set of problems that are now gaining attention. Instead of asking the question "How does one organism work?" we are now interested in "Who is here in this sample and what are they doing?". Since greater than 99% of microbes cannot be cultivated in isolation [3], metagenomics is a necessity if we wish to understand the microbial diversity of our planet.

Examples of metagenomic applications include human health, soil fertility and forensics. The National Institute of Health has created an initiative called The Human Microbiome Project to examine microbes associated with health in several areas of the human body [2]. For example it is hypothesized that the human gastrointestinal tract contains microbes that outnumber human cells 10 to 1 [2]. Many of these microbes are believed to be involved with the digestive process. Most of these microbes cannot be isolated in the laboratory. Therefore they cannot be cultured for abundance so that their DNA can be extracted and amplified for genomic analysis. Instead we turn to metagenomics where we obtain the DNA of the environmental sample, extract and amplify the DNA, sequence the samples, assemble the samples and finally attempt to annotate the sequences. Annotation is certainly an elusive task since we do not know which microbes are in the sample to begin with. So we turn to sequence alignment tools such as BLAST [4, 5] which aid us in answering a fundamental question in metagenomics, namely "Who is here?". Before we can fully trust the results of BLAST for taxonomic classification, we seek to benchmark how database representation affects its performance.

2. Background on Taxonomy

Answering the question "Who is here?" is an issue of taxonomy. Taxonomy refers to the science of naming and classifying organisms. The National Center for Biotechnology Information (NCBI) maintains a taxonomy database which is considered a well respected source by the scientific community for taxonomic information [4]. The standard hierarchy of the taxonomy used in this paper is Phyla, Order, Family, Genus, Species, Strains as recommended by the NCBI. As of September 2009 there are over 339,500 taxa represented in the database. Of these taxa 968 are completely sequenced genomes of microbial organisms. Clearly, this is only a small fraction of the microbes inhabiting our planet today, however, the databases are expanding rapidly and as the field of metagenomics becomes more pervasive we shall see substantial increases in the number of taxa maintained in these databases.

When an organism's DNA or metagenomic sample has been sequenced it is a natural step to compare this new sequence to existing, annotated sequences in the databases for similarity [6, 7]. BLAST (Basic Local Alignment Search Tool) is both a web based and standalone tool developed by the NCBI for comparing sequence similarity

between two nucleotide or protein sequences [5]. The most popular way researchers use the tool is to input a sequence as a query against the public sequence databases which include NCBI Taxonomy (http://www.ncbi.nlm.nih.gov/Taxonomy/). BLAST returns sequences that are similar to the input query. BLAST will attempt to align the query with the sequences in the databases and then issue a statistical report to provide a level of confidence in the alignment. BLAST is actively maintained by the NCBI and can be found here (http://blast.ncbi.nlm.nih.gov/Blast.cgi).

The first alignment in the report returned by BLAST is supposedly the sequence in the database with the greatest similarity to the query sequence. When the query sequence is small (e.g. < 500bp), BLAST tends to produce multiple ambiguous top-hits. It has been found that the closest BLAST hit is often not the nearest-neighbor [8]. Generally speaking, microbiologists rely on the BLAST results without question [9, 10, 11]. Researchers have now begun to analyze and compare the performance of BLAST for metagenomic datasets. The findings are indicating that classifying genome sequence fragments based on the best BLAST hit only yield reliable results if there are close relatives represented in database for comparison [12, 13].

3. Method

A total of 635 distinct microbial strains downloaded in 2008 from the NCBI Genbank database were considered for our experiments. We have found that each of the 635 strains in our database can be classified to one of 19 different phyla and 272 different genera. In order to partition the database for our experiments we decided to focus on two well-represented and two under-represented classes each for the levels of phyla and genus. Thus two separate experiments were performed; one for the level of phyla and the other for genus. Table 1 shows the composition of each class for each experiment.

Table 1. The class composition for the phyla and genus five-fold cross validation experiments are provided below. A total of 463 strains were included in the phyla experiment. We chose to use two phyla having well-representation and two having under-representation in the database. For example, Proteobacteria (well) accounted for 315 (68%) of the 463 strains included in the experiment. These strains were partitioned into five groups each containing 63 strains. The remaining three classes were partitioned in the same manner ensuring that approximately 20% of the strains belonging to the class were in each group. The first group from all four classes was combined and BLAST against the remaining four groups. This procedure was repeated five times so that each group was used for query once. An identical procedure was used at for the genus experiment.

Phyla					
Total Strains – 463		Database (80%) – 370		Query (20%) – 93	
Well-Represented			Under-Represented		
Class	# of Strains	# Queries Sampled	Class	# of Strains	# Queries Sampled
Proteobacteria (well1)	315 (68%)	63	Crenarchaeota (under1)	15 (3%)	3
Fermicutes (well2)	116 (25%)	23	Tenericutes (under2)	17 (4%)	4
Genus					
Total Strains – 64		Database (80%) – 51		Query (20%) – 13	
Well-Represented			Under-Represented		
Class	# of Strains	# Queries Sampled	Class	# of Strains	# Queries Sampled
Steptococcus (well1)	26 (40%)	5	Yersinia (under1)	10 (16%)	2
Staphylococcus (well2)	18 (28%)	4	Synechococcus (under2)	10 (16%)	2

The two well-represented classes were chosen to be the two classes at each level that contained the greatest amount of microbial strains. For example, the phyla class Proteobacteria contained 315 strains out of the 635 strains in the overall database. The two under-represented classes were chosen arbitrarily so that they each contain no more than 20 strains. Many classes in the database contained only 1 strain; however the five-fold cross validation statistical measure necessarily requires that we have a minimum of 5 strains. We chose under-represented classes containing 10 to 17 strains as shown in Table 1.

The five-fold cross validation experiments proceeded as the following for phyla using 500bp query fragments. The identical procedure was followed for the level of genus thus yielding two separate experiments. The distribution

of the classes for the experiments can be found in Table 1. To measure the possible effect of class bias, we also performed an equal class representation experiment at the phyla level as discussed in the results section.

Since we have chosen five-fold cross validation, we randomly partitioned the strains from each class into five groups. The first group from each class was combined to create a set of query strains. To simulate a metagenomics dataset obtained using the next generation of 454 pyrosequencing technology [14], each query strain's genome was randomly sampled extracting 100 fragments each 500bp in length. Each fragment was annotated with its membership class so that we could determine if BLAST correctly matched the fragment. These sampled fragments were used as queries for BLAST sequence alignment. The whole-genomes of the remaining strains were used to construct the BLAST training database in which BLAST would attempt to align against the query sequences. For example, in the phyla experiment, 93x100 (20%) query fragments were BLAST against a database of 370 (80%) whole-genomes comprised of the remaining strains belonging to the 4 phyla. The percent accuracy is calculated as the number of query fragments correctly identified by BLAST over the total number of query fragments. This procedure was repeated a total of five repetitions so that each strain was in the query test set once. The results from the five partitions were averaged and the standard deviation was calculated. A survey of cross validation methods can be found from these sources [15, 16, 17].

BLAST may potentially return multiple ambiguous hits meaning that all of the top scores returned have the same statistical expect value (e-value). In these instances all of the aligned sequences must be from the true taxonomic class otherwise the BLAST result was marked incorrect for the corresponding query sequence. Additionally, BLAST may not return a report for a query sequence that it has determined to be a low-complexity region. In these few instances we marked the query as incorrect. While this filter may be turned off we've found that BLAST consumes significantly more resources; therefore we've chosen to leave it in the default setting.

4. Results

4.1 Well/Under Representation Experiments

The results of the two cross-validation experiments with well/under representation are summarized in Table 2. Each experiment had four classes; two classes that were well-represented by strains in the dataset and two classes that were under-represented. The percent accuracy is the number of strain fragments that BLAST matched with the correct class over the total number of fragments. The average score reported is the average of all five repetitions of the cross validation experiment. The standard deviation is calculated in a similar manner. Individual scores for each repetition, for all experiments are provided in Appendix A.1.

In addition to the percent accuracy of BLAST across all strains for each experiment, Table 2 lists the accuracy of BLAST on the four individual classes as well as the accuracy on the combined well and under represented classes. Each of these combined groups contains two classes.

Table 2. The percent accuracy scores of BLAST for the genus and phyla experiments are provided below. BLAST was marked correct if it matched the query fragment to the correct class and incorrect otherwise. It was also marked incorrect if it provided multiple ambiguous hits whereupon these hits belonged to two or more different classes. The percent accuracy for each cross validation repetition is the number of correct matches over the total number of query fragments. The percent accuracy scores over all five repetitions were average and are provided below along with the standard deviation of scores.

Percent Accuracy		All	Well	Well 1	Well 2	Under	Under 1	Under 2
500bp Genus	AVG	95.87	96.60	95.65	97.87	94.15	96.90	91.40
	STD	2.10	3.10	4.91	3.03	3.57	4.56	8.51
500bp Phyla	AVG	87.21	90.06	92.67	83.01	48.74	36.80	59.38
	STD	2.29	2.30	0.79	7.80	9.64	16.43	14.52

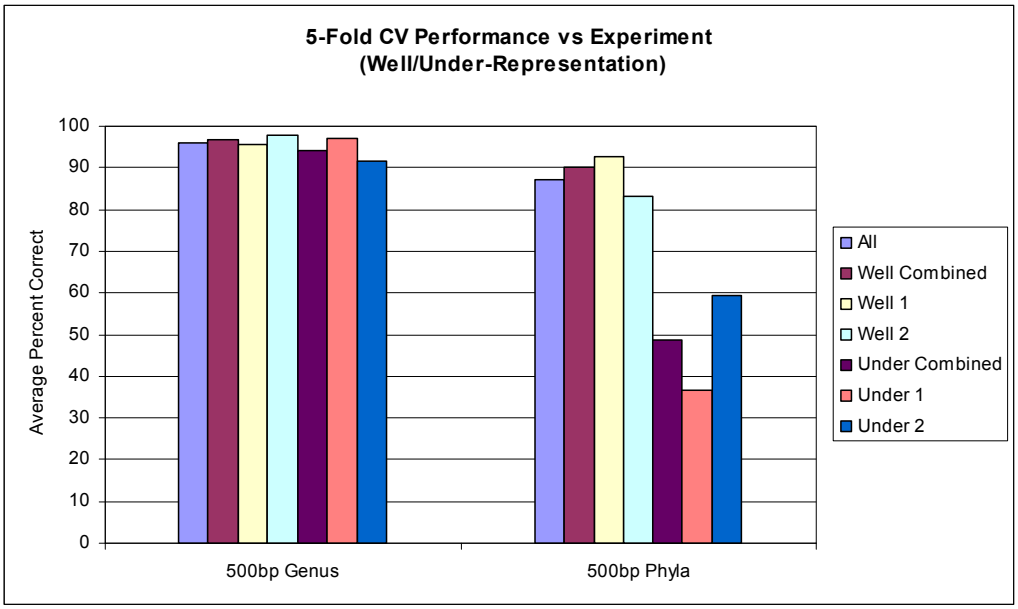

Figure 1. This bar graph illustrates the data provided in Table 2. All four classes in the genus experiment exhibited similar percent accuracy scores. However, there is a clear difference in percent accuracy between the well- and under-represented classes in the phyla experiment. We've found that this is due in part to the genus level having less diversity than the phyla level.

All seven different scores for percent accuracy are plotted against the two experiments in Figure 1. The trend across the two experiments indicates that accuracy increasingly diminishes moving from 500bp genus to 500bp phyla. It is evident from Figure 1 that the percent accuracy of all strains for each experiment is highly dependent on the accuracy of BLAST correctly identifying the fragments belonging to strains having membership in the under-represented classes. This is evident from the disparity between the under-represented scores for genus and phyla. Genus under-represented has an accuracy nearly 40% higher than the phyla under-represented group. Predictably, genus percent accuracy for all strains is nearly 10% higher than phyla.

4.1.1 Phyla

There were a total of 463 strains considered in the phyla experiments. Each cross-validation repetition consisted of 93 (20%) strains chosen at random without replacement to BLAST against the remaining 370 (80%) strains in the dataset (see Table 1). Accordingly, every strain in the dataset was used as a test strain once. Each test strain was sampled randomly 100 times; each sample consisted of a fragment 500bp in length. These 100 fragments were used in place of the test strain. The phyla experiment shows that well-represented strains scored approximately 40% higher than under-represented strains.

Strains belonging to the under-represented class Crenarchaeota were misclassified 78% on average. These misclassified fragments were frequently matched with strains belonging to the well-represented phyla. For example, BLAST classified 74% of fragments belonging to *Pyrobaculum aerophilum* to Proteobacteria (well) rather than Crenarchaeota (under) as expected.

In general, when BLAST misclassified fragments, 5% of the misclassifications belonged to a strain in the under-represented classes. Of the remaining 95% misclassified fragments, approximately 72% of the misclassifications went to strains belonging Proteobacteria (well) alone. This is similar to chance since the database is comprised of 93.3% well-represented strains. Furthermore, Proteobacteria (well) makes up for about 73% of the combined well-represented classes. Generally speaking, BLAST frequently confused under-represented fragments with well-represented phyla.

4.1.2 Genus

There were a total of 64 strains considered in the genus experiments. Each cross-validation repetition consisted of 13 (20%) strains chosen at random without replacement to BLAST against the remaining 51 (80%) strains in the dataset (see Table 1). Accordingly, every strain in the dataset was used as a test strain once. Each test strain was sampled randomly 100 times; each sample consisted of a fragment 500bp in length. The genus experiment shows that well-represented strains scored marginally higher than under-represented strains.

BLAST misclassified 25% of *Synechococcus CC9311* (under) 500bp fragments with strains belonging to the other three genera. 76% of these misclassifications went to a well-represented genus. When BLAST misclassified a fragment 21% of the misclassifications belonged to strains in the under-represented classes. Of the remaining 79% misclassified fragments, 48% went to strains belonging to the well-represented to Staphylococcus (well) alone. Generally speaking, BLAST frequently confused under-represented fragments with well-represented genus.

4.2 Equal Representation Experiments

The result of the phyla cross validation experiment with equal class representation is presented in Table 3. A dataset consisting of 60 strains was constructed to have equal representation among the four phyla classes. 15 strains were randomly sampled from each class from our original phyla dataset of 463 strains. Each cross-validation repetition consisted of 12 (20%) strains chosen at random without replacement to BLAST against the remaining 48 (80%) strains in the dataset (see Appendix A.2). Accordingly, every strain in the dataset was used as a test strain once. Each test strain was sampled randomly 100 times; each sample consisted of a fragment 500bp in length. These 100 fragments were used in place of the test strain.

Table 3. The result of the equal-representation cross validation experiment is provided below. The first score column is the percent accuracy of BLAST for all four classes while the four columns to the right, labeled with the phyla's abbreviated name, refer to the individual scores for each class considered in this experiment. These are the same four phyla considered in the well/under represented experiment except now we have selected 15 strains from each class for this experiment so that each repetition will have equal-representation among the four classes in the database. The cross validation procedure used here is identical to the one used for the well/under experiments.

Percentages		All	Prot	Firm	Cren	Ten
500bp Phyla	AVG	63.75	73.33	61.53	53.46	66.67
	STD	8.23	13.88	10.35	10.78	13.44

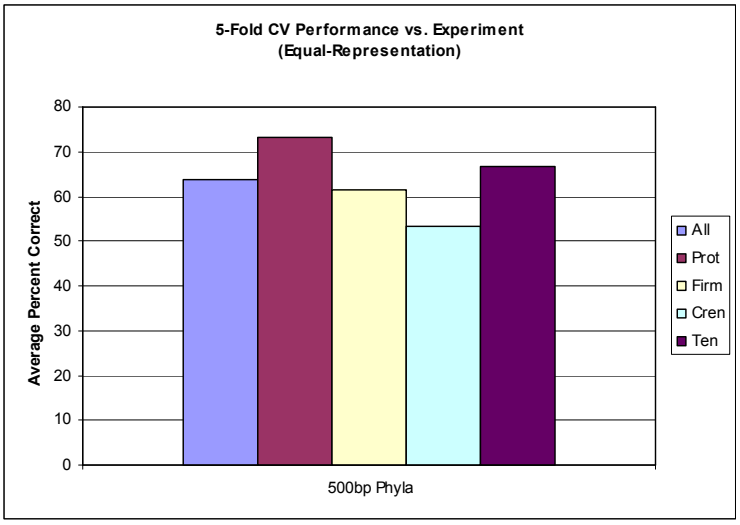

Figure 2. This bar graph illustrates the data provided in Table 3. The four classes in the equal-representation experiment have percent accuracy performance much more similar to one another than in the well/under representation experiment. This finding indicates that class composition in the database affects the performance of BLAST with well-represented classes having higher accuracy than under-represented classes.

The average percent accuracy for the experiment is shown in Figure 2. The overall percent accuracy decreased by ~25% from the well/under represented phyla experiment while the standard deviation increased from ~2.5% to ~8%. The percent accuracy decrease is mostly contributed to the decrease in accuracy (~29%) of Proteobacteria. The increase in standard deviation is also most significantly affected by Proteobacteria whose standard deviation increased by ~14%. This increase was expected since Proteobacteria was considered a well-represented class in the previous experiments and we've found that variance decreases with increasing representation.

5. Discussion

It is clear from our experiments that the accuracy of BLAST is highly dependent on the composition of the training database. The well/under phyla experiment confirmed that the well-represented classes have nearly 40% higher accuracy than the under-represented classes. Still BLAST is performing much better than chance on all classes. For phyla (500bp) we see that Proteobacteria (well) scored 92.67%. With a database composition of 252/370 we confirm that this score is much higher than chance which would be about 68%. This can also be verified for under-represented classes. For instance BLAST scored 59.38% for Tenericutes (under). Given its database composition we would expect a percent accuracy of 14/370 or 3.7% by chance.

Upon further examination of the database's composition we observe the ratio of well to under-represented strains in the phyla database is nearly 14:1 (Figure 3). Incidentally, by chance, if we rolled a die we would expect BLAST to classify a strain to a well-represented class 14/15 or 93.3% percent of the time. While we found through our experiments that BLAST is able to classify much better than chance, the allocation of BLAST misclassifications follows a different trend. For example, in the phyla experiments when BLAST misclassified a fragment ~ 95% (nearly chance) of the fragments were assigned to a well-represented class.

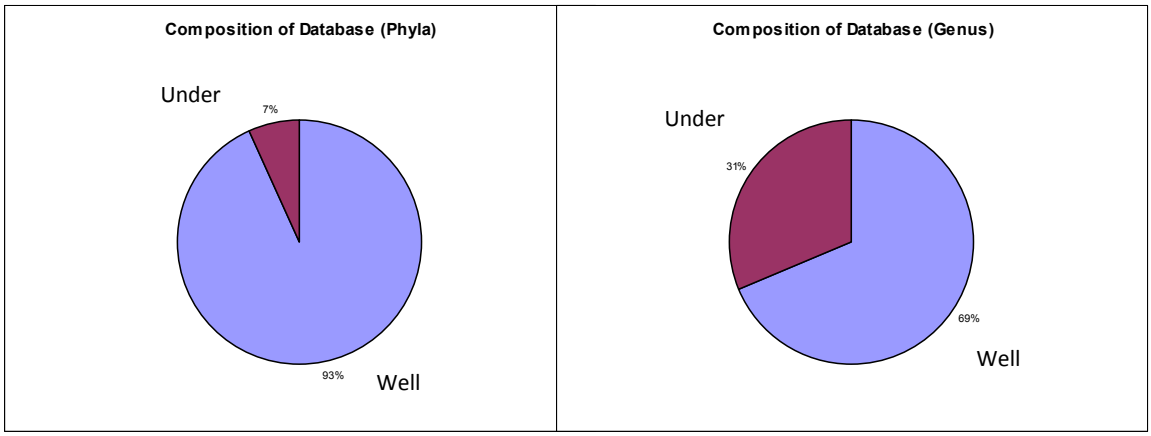

Figure 3. The pie charts above show the class composition for both the phyla and genus well/under experiments. The well-represented classes in the phyla experiment had nearly 40% higher percent accuracy scores than the under-represented classes. The difference in percent accuracy scores between the classes in the genus experiment was marginal.

These trends are also reflected in the genus experiments. For example for genus (500bp) we find that Streptococcus (well) scored 96.6%. By chance we would observe 21/51 or 41.1% accuracy. For Yersinia (under) BLAST scored 91.4% while a score by chance would be 8/51 or 15.6%. As shown in Figure 3 the genus database composition is about 2.2:1 predicting that BLAST would classify a strain to a well-represented class about 69% of the time by chance. This is reflected in the allocation of BLAST misclassifications where about 76% of the BLAST misclassified fragments went to a well-represented class.

The chart in Figure 1 indicates that the genus experiment outperformed the phyla experiment. We hypothesized that this was due to the difference in the well/under database composition (Figure 3) between the sets of experiments. Since the phyla under-represented composition is only 7% of the entire database as opposed to 31% for

the genus experiments, we wanted to find out if phyla overall percent accuracy score would approach the scores for the genus experiments if we increased the percent composition of phyla's under-represented classes.

We conducted the equal-representation experiment among the four classes at the phyla level to further test for compositional bias. The results show that the scores for the two under-represented classes increased while the scores for the two well-represented classes decreased substantially resulting in similar performance for all 4 phyla classes. Therefore we find that class composition size affects performance; improvement for the well-represented classes and degraded accuracy for under-represented classes. Proteobacteria's (well1) percent accuracy decreased 29% while Firmicutes (well2) decreased 22%. We infer that strains belonging to the genus level have less diversity than at the phyla level since there was such a substantial decrease in scores at the phyla level when the dataset was reduced to the size of the under-represented genus classes. There is proof to show that 16s rRNA sequences have 6% divergence at the genus level and 3% for species so we infer that this percentage is higher at the phyla level [18].

The standard deviation for classes in the phyla well/under represented experiment is shown in Figure 4. It is clear that class size has a significant effect on the standard deviation of percent accuracy scores. Proteobacteria with 252 strains had a standard deviation of 0.79% while Crenarchaeota with 12 strains had a standard deviation of 16.43%. The overall standard deviation for phyla increased from 2.29% to 8.23% moving from the well/under represented to the equally represented experiment. Database composition remains an important consideration in BLAST experiments in addition to the certainly of class labels.

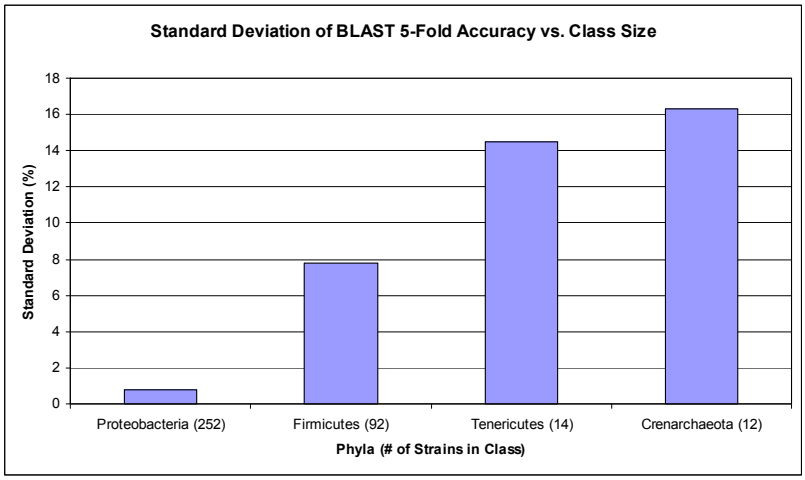

Figure 4. The bar graph above shows the standard deviations of BLAST's percent accuracy for each class over all five repetitions in the phyla well/under represented experiment. This graph clearly shows that for phyla the standard deviation of percent accuracy scores reported from BLAST decreases with increasing examples in the database.

6. Conclusion

Several five-fold cross validation experiments were examined in this study. Our analysis has shown that database composition contributes substantially to the accuracy of the BLAST algorithm. Overall we've found that the standard deviation of percent accuracy scores decreases with increasing class representation in the database. We've also demonstrated that BLAST performs much better than chance even when representation in the database is much smaller than the majority as shown in the phyla experiments. However, when BLAST misclassifies a fragment, it appears to assign the fragment by chance.

As shown in Figure 1 the genus experiment scored higher than the phyla experiment. We've shown through our equal-representation phyla experiment that this marked difference is most likely attributed to the genus level having better definition and less diversity than the phyla level. The study demonstrates the intuitive result that a user of BLAST would want to build a database having as much representation as possible to increase accuracy and decrease the standard deviation of scores. As the numbers from the experiments show, the output of BLAST is clearly not always the correct match and we suggest the use of additional, external information to form a consensus of matches.

Acknowledgments

The work in this paper was supported by the National Science Foundation CAREER award #0845827.

Appendix

A.1 CV Results - Well/Under Representations

Table 4. Results of the well/under cross validation experiment at the genus level, 500bp fragments

Genus Results – 500bp			
Overall Dataset			
Repetition	Training	Test	% Correct
1	51	13	97.15
2	51	13	93.23
3	51	13	98.62
4	52	12	94.67
5	51	13	95.69
		AVG	**95.87**
		STD	**2.10**

Genus Well-Represented				**Genus Under-Represented**			
Combined				**Combined**			
Repetition	# Training	# Test	% Correct	Repetition	# Training	# Test	% Correct
1	35	9	99.00	1	16	4	93.00
2	35	9	93.44	2	16	4	92.75
3	35	9	99.11	3	16	4	97.50
4	36	8	93.00	4	16	4	98.00
5	35	9	98.44	5	16	4	89.50
		AVG	**96.60**			**AVG**	**94.15**
		STD	**3.10**			**STD**	**3.57**
Streptococcus				**Yersinia**			
Repetition	# Training	# Test	% Correct	Repetition	# Training	# Test	% Correct
1	21	5	100.00	1	8	2	100.00
2	21	5	88.20	2	8	2	89.00
3	21	5	99.00	3	8	2	97.50
4	21	5	93.20	4	8	2	98.00
5	20	6	97.83	5	8	2	100.00
		AVG	**95.65**			**AVG**	**96.90**
		STD	**4.91**			**STD**	**4.56**
Staphylococcus				**Synechococcus**			
Repetition	# Training	# Test	% Correct	Repetition	# Training	# Test	% Correct
1	14	4	97.75	1	8	2	86.00
2	14	4	100.00	2	8	2	96.50
3	14	4	99.25	3	8	2	97.50
4	15	3	92.67	4	8	2	98.00
5	15	3	99.67	5	8	2	79.00
		AVG	**97.87**			**AVG**	**91.40**
		STD	**3.03**			**STD**	**8.51**

Table 5. Results of the well/under cross validation experiment at the phyla level, 500bp fragments

Phyla Results – 500bp			
Overall Dataset			
Repetition	Training	Test	% Correct
1	370	93	88.51
2	370	93	89.94
3	371	92	84.05
4	371	92	87.62
5	370	93	85.94
		AVG	**87.21**
		STD	**2.29**

Phyla Well-Represented				**Phyla Under-Represented**			
Combined				**Combined**			
Repetition	# Training	# Test	% Correct	Repetition	# Training	# Test	% Correct
1	345	86	90.98	1	25	7	58.14
2	345	86	93.47	2	25	7	46.57
3	345	86	87.50	3	26	6	34.67
4	345	86	89.72	4	26	6	57.50
5	344	87	88.63	5	26	6	46.83
		AVG	**90.06**			**AVG**	**48.74**
		STD	**2.30**			**STD**	**9.64**
Proteobacteria				**Crenarchaeota**			
Repetition	# Training	# Test	% Correct	Repetition	# Training	# Test	% Correct
1	252	63	91.94	1	12	3	50.67
2	252	63	93.29	2	12	3	28.00
3	252	63	91.71	3	12	3	14.67
4	252	63	93.44	4	12	3	36.00
5	252	63	92.97	5	12	3	54.67
		AVG	**92.67**			**AVG**	**36.80**
		STD	**0.79**			**STD**	**16.43**
Firmicutes				**Tenericutes**			
Repetition	# Training	# Test	% Correct	Repetition	# Training	# Test	% Correct
1	93	23	88.35	1	13	4	63.75
2	93	23	93.96	2	13	4	60.50
3	93	23	75.96	3	14	3	54.67
4	93	23	79.52	4	14	3	79.00
5	92	24	77.25	5	14	3	39.00
		AVG	**83.01**			**AVG**	**59.38**
		STD	**7.80**			**STD**	**14.52**

A.2 CV Results - Equal Representation

Table 6. Results of the equal-representation cross validation experiment at the phyla level, 500bp fragments

Phyla Results – 500 bp			
Overall Dataset			
Repetition	Training	Test	% Correct
1	48	12	61.50
2	48	12	73.50
3	48	12	69.42
4	48	12	52.08
5	48	12	62.25
		AVG	**63.75**
		STD	**8.23**

Proteobacteria				**Crenarchaeota**			
Repetition	# Training	# Test	% Correct	Repetition	# Training	# Test	% Correct
1	12	3	72.00	1	12	3	45.33
2	12	3	91.33	2	12	3	71.00
3	12	3	61.67	3	12	3	56.33
4	12	3	58.67	4	12	3	45.33
5	12	3	83.00	5	12	3	49.33
		AVG	**73.33**			**AVG**	**53.46**
		STD	**13.88**			**STD**	**10.78**
Firmicutes				**Tenericutes**			
Repetition	# Training	# Test	% Correct	Repetition	# Training	# Test	% Correct
1	12	3	57.00	1	12	3	71.67
2	12	3	56.33	2	12	3	75.33
3	12	3	80.00	3	12	3	79.67
4	12	3	58.00	4	12	3	46.33
5	12	3	56.33	5	12	3	60.33
		AVG	**61.53**			**AVG**	**66.67**
		STD	**10.35**			**STD**	**13.44**

References

1. V. Kunin, A. Copeland, A. Lapidus, K. Mavromatis and P. Hugenholtz, *Micro Mol Biol Rev.* **72**, 4 (2008).
2. G. Rosen, B. Sokhansanj, R. Polikar, M. Bruns, J. Russell, E. Garbarine, S. Essinger, and N. Yok, *Current Genomics*. **10**, 7 (2009).
3. J. Handelsman, Committee on Metagenomics: Challenges and Functional Applications, N. R. Council, Ed. *The National Academies Press*, (2007).
4. S. Altschul, W. Gish, W. Miller, E. Myers, and D. Lipman, *J. Mol Biol*. **215**, 3 (1990).
5. T. Madden, *The NCBI Handbook.* Ch. 16, 1-17 (2003).
6. J. Venter, K. Remington, J. Heidelberg, A. Halpern, D. Rusch, J. Eisen, D. Wu, I. Paulsen, K. Nelson, W. Nelson, D. Fouts, S. Levy, A. Knap, M. Lomas, K. Nealson, O. White, J. Peterson, J. Hoffman, R. Parsons, H. Tillson, C. Pfannkoch, Y. Rogers, and H. Smith, *Science*. **304**, 5667 (2004).
7. M. Tress, D. Cozzetto, A. Tramontano, and A. Valencia, *BMC Bioinformatics*. **7**, 213 (2006).
8. L. Koski and G.Golding, *J. Mol Evol.* **52**, 6 (2001).
9. A. Andersson, M. Lindberg, H. Jakobsson, F. Backhed, P. Nyren, and L. Engstrand, *PLoS One.* **3**, 7 (2008).

10. D. Huson, A. Auch, J. Qi, and S. C. Schuster, *Genome Res*. **17**, 3 (2007).
11. S. Havre, B. Webb-Roberston, A. Shah, C. Posse, B. Gopalan, and F. Brockman, *Proceedings of the 2005 IEEE Computational Systems Bioinformatics Conference*, 341–350 (2005).
12. K. E. Wommack, J. Bhavsar, and J. Ravel, *Appl Environ Microbiol.* **74**, 5 (2008).
13. C. Manichanh, C. Chapple, L. Franguel, K. Gloux, R. Guigo and J. Dore, *Nucleic Acids Res*. **36**, 16 (2008).
14. L. Krause, N. Diaz, A. Goesmann, S. Kelley, T. Nattkemper, F. Rohwer, R. Edwards, and J. Stoye, *Nucleic Acids Res*. **36**, 7 (2008).
15. G. L. Rosen, E. M. Garbarine, D. A. Caseiro, R. Polikar, and B. A. Sokhansanj, *Hindawi Adv Bioinfo*. **2008**, (2008).
16. R. Kohavi, *Proceedings of Fourteenth International Joint Conference on Artificial Intelligence (IJCAI)*, 1137–1143 (1995).
17. P. Devijver and J. Kittler. Pattern Recognition: A Statistical Approach, *Prentice-Hall*, London. (1982).
18. J. Clarridge, *Clin Microbiol Rev*. **17**, 4 (2004).