

Results and Discussion

Analysis and statistics from the CONSORF evaluation results

The accuracy of predicted CDSs in terms of sensitivity and specificity varies from organism to organism. Since start site prediction is more difficult than stop site prediction, owing to numerous theoretically plausible starts for each stop, the evaluation was performed on four different levels based on the strictness of the start prediction (Table 1). For more strict evaluation regarding authentic coding genes, both the annotated and the overall data sets of 'public CDSs' were used as a reference set of evaluation. The annotated and the overall 'public CDSs' exclude and include hypothetical CDSs (CDSs with no functional annotation), respectively.

Using the stop position match only (level 1 on the CONSORF evaluation ladder), the average sensitivities of CONSORF ('representative CDSs') were 99.1% and 95.0% in the comparison with the annotated and the overall 'public CDSs', respectively. In addition, the average specificities were 66.7% (annotated 'public CDSs') and 95.5% (overall 'public CDSs'), respectively. The low specificity in the comparison with the annotated 'public CDSs' is due to the relatively excessive number of predicted 'representative CDSs' compared to the annotated 'public CDSs'. When stop and any candidate start sites were used (level 2), average sensitivities and specificities were 95.3% and 64.1% (annotated 'public CDSs') and 90.7% and 91.2% (overall 'public

CDSs'), respectively. For the same CDSs, based on the shortest start (level4), they were 77.6% and 52.2% (annotated 'public CDSs') and 72.0% and 72.4% (overall 'public CDSs'), respectively.

Highly-confident homology CDSs

To improve the low specificity and to increase the overall accuracy (F-measure) of CONSORF in the comparison with the annotated 'public CDSs', CONSORF provides 'highly-confident homology CDSs'. They are a subset of 'homology CDSs' with a high reliability score (a sum of bit scores from pair-wise genome-to-proteome sequence comparison) over a given cutoff. In the comparison with the annotated 'public CDSs', their average sensitivity and specificity were 90.1% and 80.7%, respectively. Since 'homology CDSs' and 'ab initio CDSs' have reliability scores, it is possible to increase their specificities at the cost of decreasing sensitivities by adjusting their reliability cutoff scores. In addition, since consensus-based 'homology CDSs' showed a slightly better accuracy than the individually best 'alternative CDSs', the former were used to provide 'highly-confident homology CDSs' as well as 'integrated CDSs'.

High-accuracy and biased CDS predictions

As expected, organisms that have any close neighbors in the comparison, such as *Staphylococcus aureus*, *Chlamydomonas reinhardtii*, and *Listeria monocytogenes*, showed very high sensitivity and specificity by CONSORF. However, in some organisms such as *Xanthomonas oryzae* KACC10331 [24], *Mycoplasma gallisepticum*

[25], *Rhodospirellula baltica* [26], and *Mycobacterium leprae* [27], the accuracy of CONSORF was not satisfactorily high. We investigated these cases and were able to find a systematic bias in the original prediction of ‘public CDSs’ or unusual biological features, such as an abundance of pseudo-genes, by analyzing the pattern of sensitivity and specificity based on four evaluation levels of CONSORF. This indicates CONSORF can detect some systematic errors in the initial CDS predictions.

Improvement of CONSORF accuracy

We estimate that the accuracy of CONSORF will increase as the quality and quantity of public databases rise. Moreover, we expect that further revisions on candidate starts will improve the accuracy of start site prediction.