APPENDIX S4. Comparison of transcript assembly methods. There are a growing number of de novo transcriptome assembly algorithms tailored specifically to RNASeq data (Velvet, Mira, transAbyss, SOAP de novo, Trinity, and others); this trend warrants a comparative approach to validate assembly quality. To this end, we selected Velvet and Trinity for comparison. Velvet is a well-cited application originally designed for short-read genome assembly (and augmented for transcriptome assembly via Oases), and Trinity is a more recently published program built specifically to assemble de novo transcriptomes from short-read sequence data. Trinity, optimized for large volumes of RNA-Seq data, analyzes clustered de Bruijn graphs in parallel, and attempts to recognize alternative splicing isoforms and paralogous genes. Velvet also makes use of de Bruijn graphs and employs a graph error correction algorithm; Oases uses Velvet's output to heuristically identify splice variants and reduce the number of declared contigs.

Using both programs, we ran five test assemblies of random paired-end read selections (5, 10, 20, 25, and 30 million pairs) and two additional full assemblies of indexed paired-end reads (Index 2 and a full-sample Index 2+4). All reads were drawn from the Corvallis sample. A comparison of these assembly results led us to select Velvet Oases as our preferred method for assembling the *B. sylvaticum* transcriptome. In comparison to Trinity, Velvet consistently produced longer average contig lengths, especially in the traditionally gene-rich 1000–1500 bp range, due primairly to Oases' ability to identify and reduce non-splicing isoforms to clusters of like contigs (Fig. D). Trinity consistently produced a large set of short contigs (<500 bp) that suppressed the average contig length and increased the likelihood of mis-assembly, given the standard trimmed read length of 100 bp (Figs. A and B). Both methods produced similar maximum contig lengths (Fig. C).