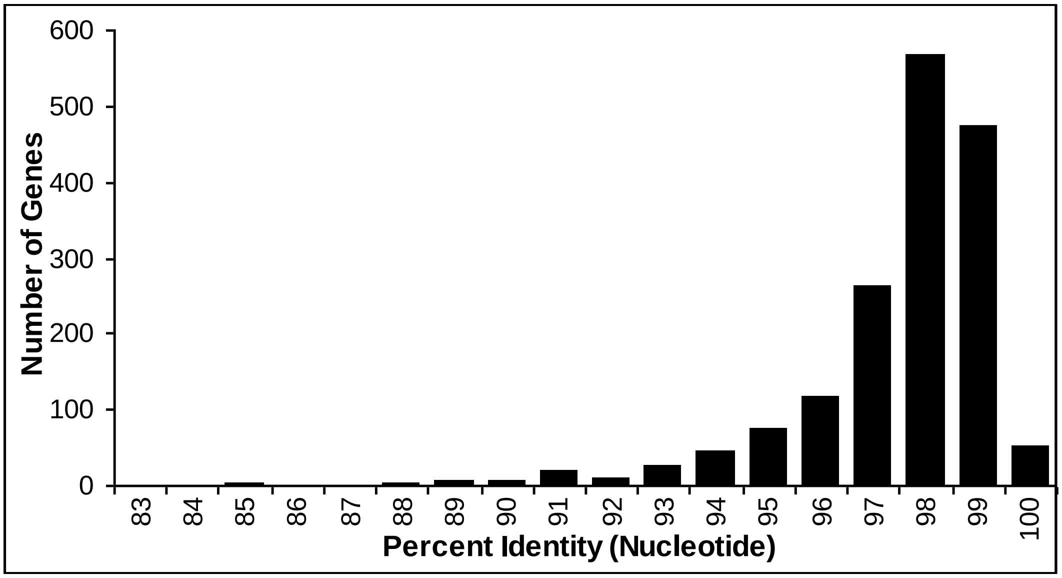
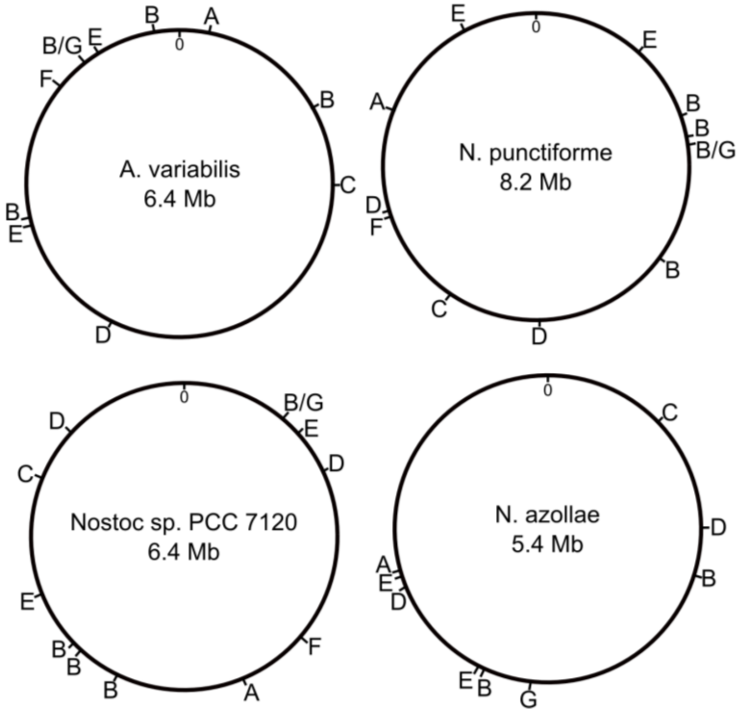
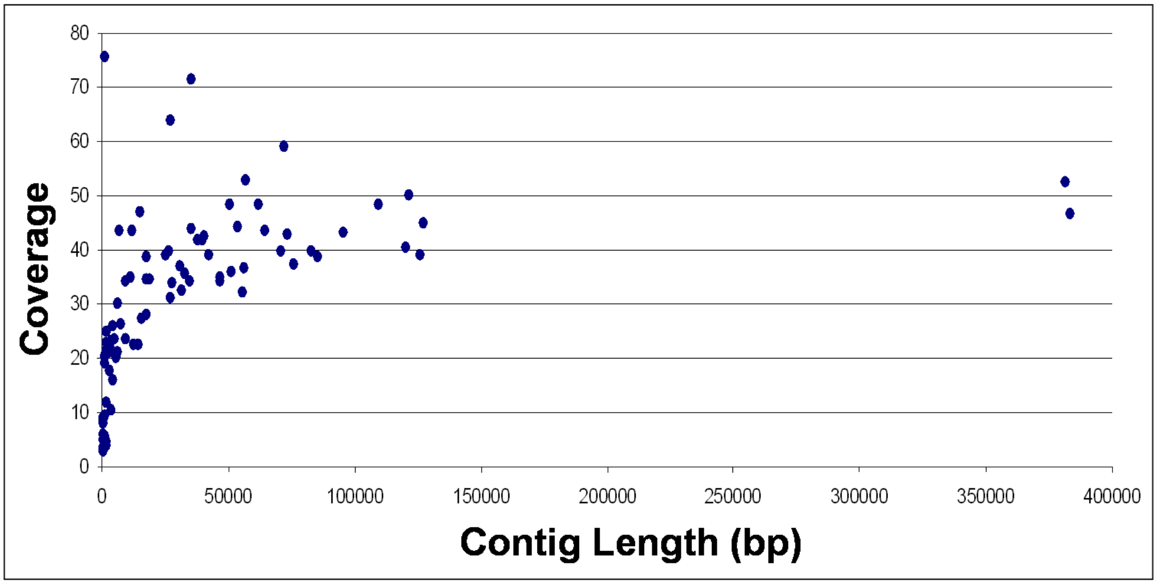


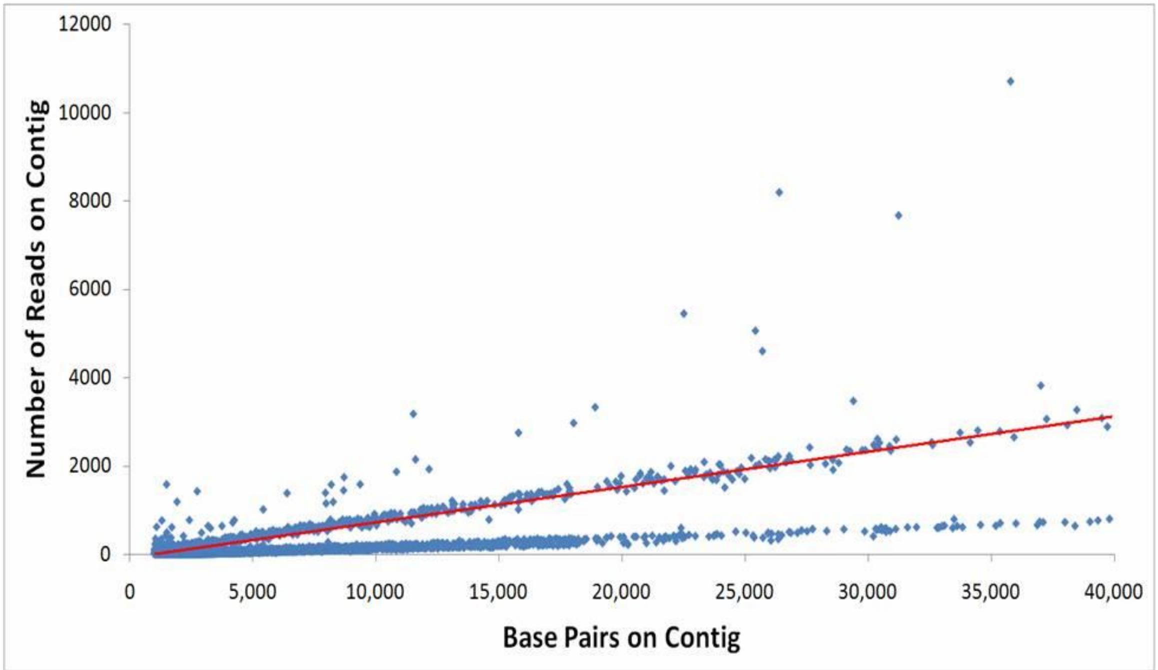
Supplementary Figure S1. RintHH and RintHM gene similarity. The percent identity at the nucleotide level of the 1,671 shared genes of *R. intracellularis* HM01 and HH01.



Supplementary Figure S2. Nitrogen metabolism genes in Nostocales genomes. The location of nitrogen metabolism-related genes across the four closed Nostocales genomes. A single letter may represent multiple genes (i.e. multiple adjacent nitrate transporters). A – GS inactivating factor 7, B – nitrate transporters, C – Fd-GOGAT, D – ammonium transporters, E – Urease subunits, F – Urea transporters, G – nitrate and nitrite reductase.



Supplementary Figure S3. RintHH assembly. The length and coverage depth of each of the 90 contigs that make up the *Richelia intracellularis* HH01 genome assembly.



Supplementary Figure S4. *Calothrix rhizosoelniae* SC01 assembly. Data from the initial assembly of 454 reads from *C. rhizosoleniae* SC01 samples (69,919 contigs, 81.4 Mb). For each contig, the number of reads assembled to each contig plotted against the length of the contig. The contigs lying along the red line (representing 30x coverage depth) are the ones presumed to be from *C. rhizosoleniae* SC01, based on blast results.

Supplementary Table S1. Obligate symbiont comparison. Gene presence/absence in the *R. intracellularis* HH01 genome of genes uniquely present or uniquely absent in the *N. azollae* genome relative to other heterocyst-forming cyanobacterial genomes (Ran et al. 2010). *N. punctiforme* sequences are provided as representative functional gene sequences for genes not found in either obligate symbiont genome.

Gene Product	Gene symbol	*N. azollae*	*R. intracellulais*	*N. punctiforme*
uracil-DNA glycosylase	*ung*	Aazo_0465	RintHH_15290	np
dihydrofolate reductase	*folA*	Aazo_1890	np	np
thymidylate synthase	*thyA*	Aazo_5011	np	np
conserved hypothetical protein		*Aazo_3194	RintHH_13480	Npun_F1238
putative GPH family sugar transporter		*Aazo_3944	RintHH_3260	Npun_F1762
photosystem II reaction center L protein (psII 5 Kd protein)	*psbL*	*Aazo_1235	RintHH_21410	Npun_F5553
lysyl-tRNA synthetase	*lysS*	np	RintHH_10460	Npun_R5202
thymidylate synthase (FAD)	*thyX*	np	RintHH_16770	Npun_R2526
conserved hypothetical protein		*Aazo_2347	RintHH_8590	Npun_F0905
chromosomal replication initiator protein DnaA	*dnaA*	*Aazo_2080	*RintHH_4500	Npun_F0001
6-phosphofructokinase	*pfkA*	*Aazo_0224	np	Npun_R0482
putative L-cysteine/cystine lyase		*Aazo_2409	np	Npun_F1365
L-lactate dehydrogenase	*ldh*	np	np	Npun_F2517
dCTP deaminase	*dcd*	np	np	Npun_F2524
heterocyst-inhibiting signaling peptide	*patS*	np	np	Npun_R5353

np – gene is not present
* – gene present only as a pseudogene