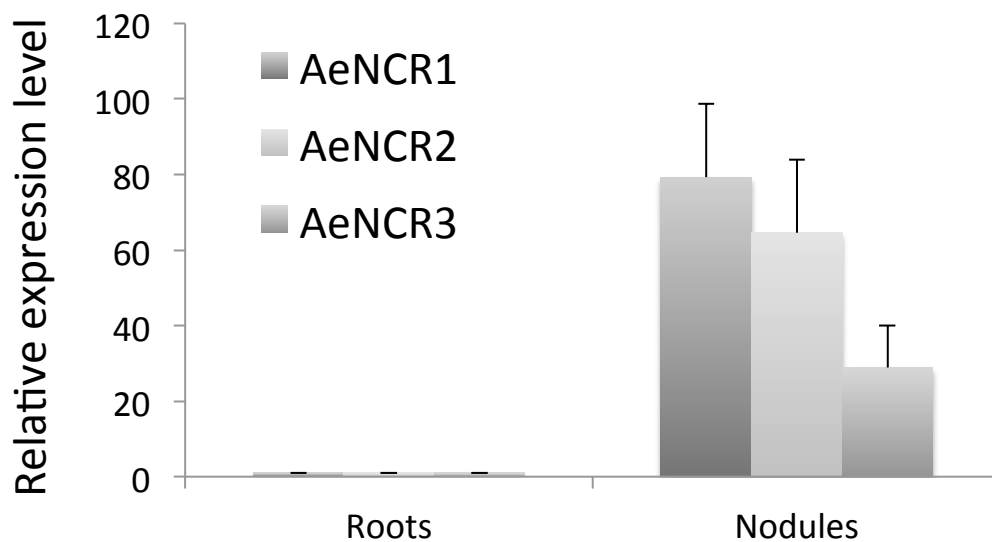
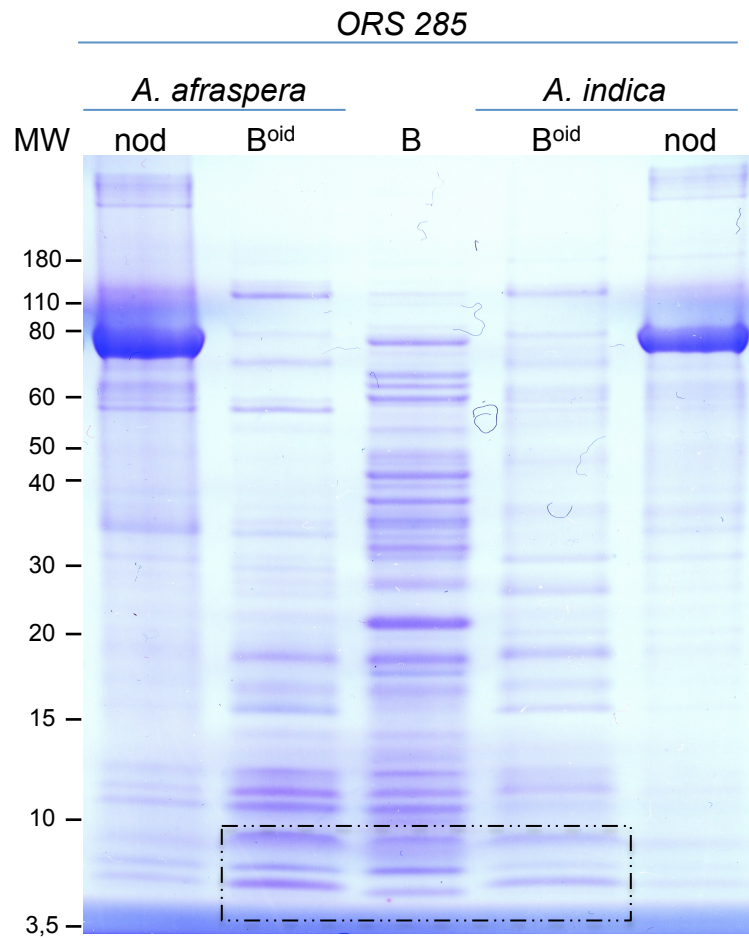


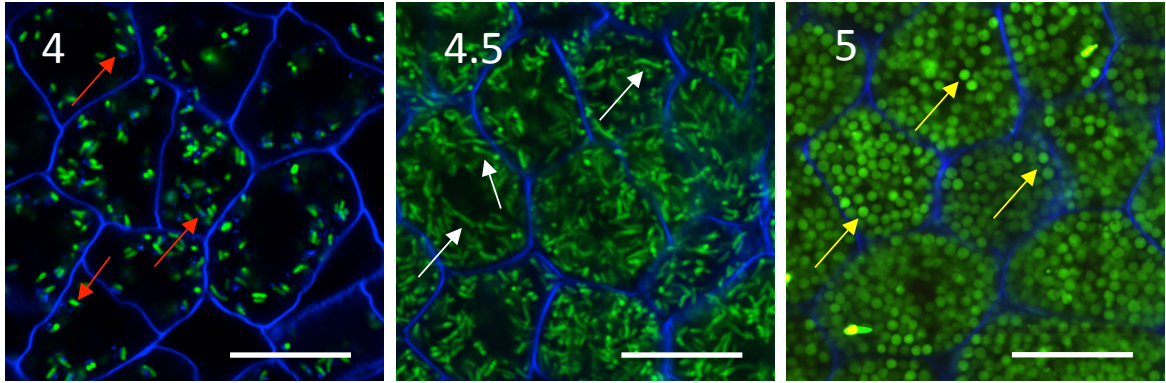
Supplemental Figure S1: Distribution of bacteroid morphotypes among papilionoid legumes. (A) Simplified phylogenetic tree with triangles representing the major legume clades. Notable species for which bacteroid morphotype has been characterized are indicated. In red, E-morphotype; green, S-morphotype and black, U-morphotype. Figure adapted from Oono et al. (2010). (B) Phylogeny of the genus *Aeschynomene* based on the ITS + *trnL*. Figure adapted from Chaintreuil et al. (2013). Species forming E-morphotype bacteroids are indicated in red, species forming S-morphotype bacteroids in green. The other species have not been analyzed.



Supplemental Figure S2. Nodule specific expression of NCR genes in *A. evenia*. RNA was extracted from *A. evenia* spp. serrulata roots or mature nodules 14 days after inoculation with the Bradyrhizobium strain ORS285. Expression levels were calculated relative to EF1 α . Error bars indicate standard errors of the means of 3 biological replicates.

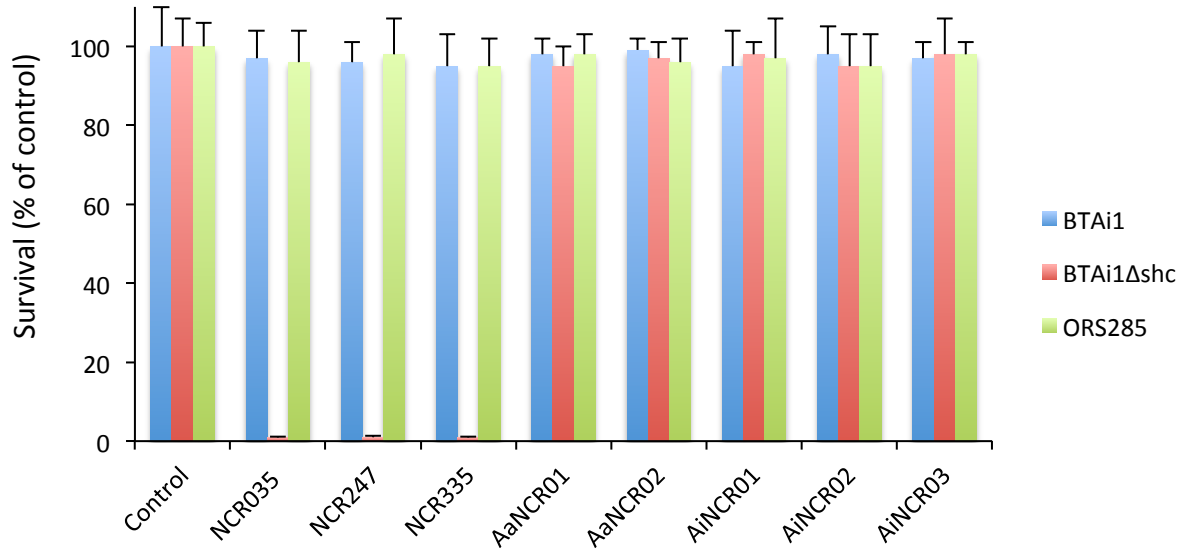


Supplemental Figure S3. NCR peptides colocalize with bacteroids in *A. afraspera* and *A. indica* nodules. Tricine-SDS-PAGE analysis and Coomassie blue staining of total protein extracts of nodules (nod), bacteroids (B^{oid}), and cultured *Bradyrhizobium* strain ORS285 cells (B) reveal the specific presence of low molecular-weight peptides in nodules and bacteroids. The rectangle in the dashed line corresponds to the region of the gel analysed by LC-MS/MS.

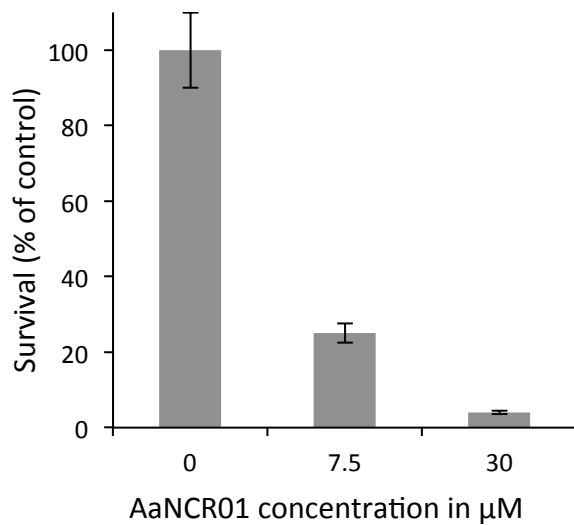


Supplemental Figure S4. Change in the bacterial cell shape during the symbiotic interaction between *Bradyrhizobium* strain ORS285 and *Aeschynomene indica*. *A. indica* nodules were elicited by a *Bradyrhizobium* ORS285 strain expressing GFP to easily visualize the bacteroids. Nodules were harvested at 4, 4.5 and 5 dpi, then sectioned and analysed by confocal microscopy. Red and yellow arrows show respectively early (U-morphotype) and final (S-morphotype) shape of bacteroids while white arrows point the transitory elongated shape of the bacteroids. Bars represent 10 µm.

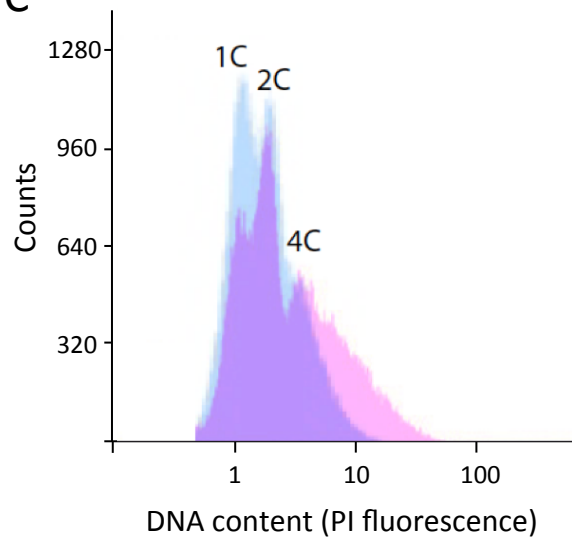
A



B



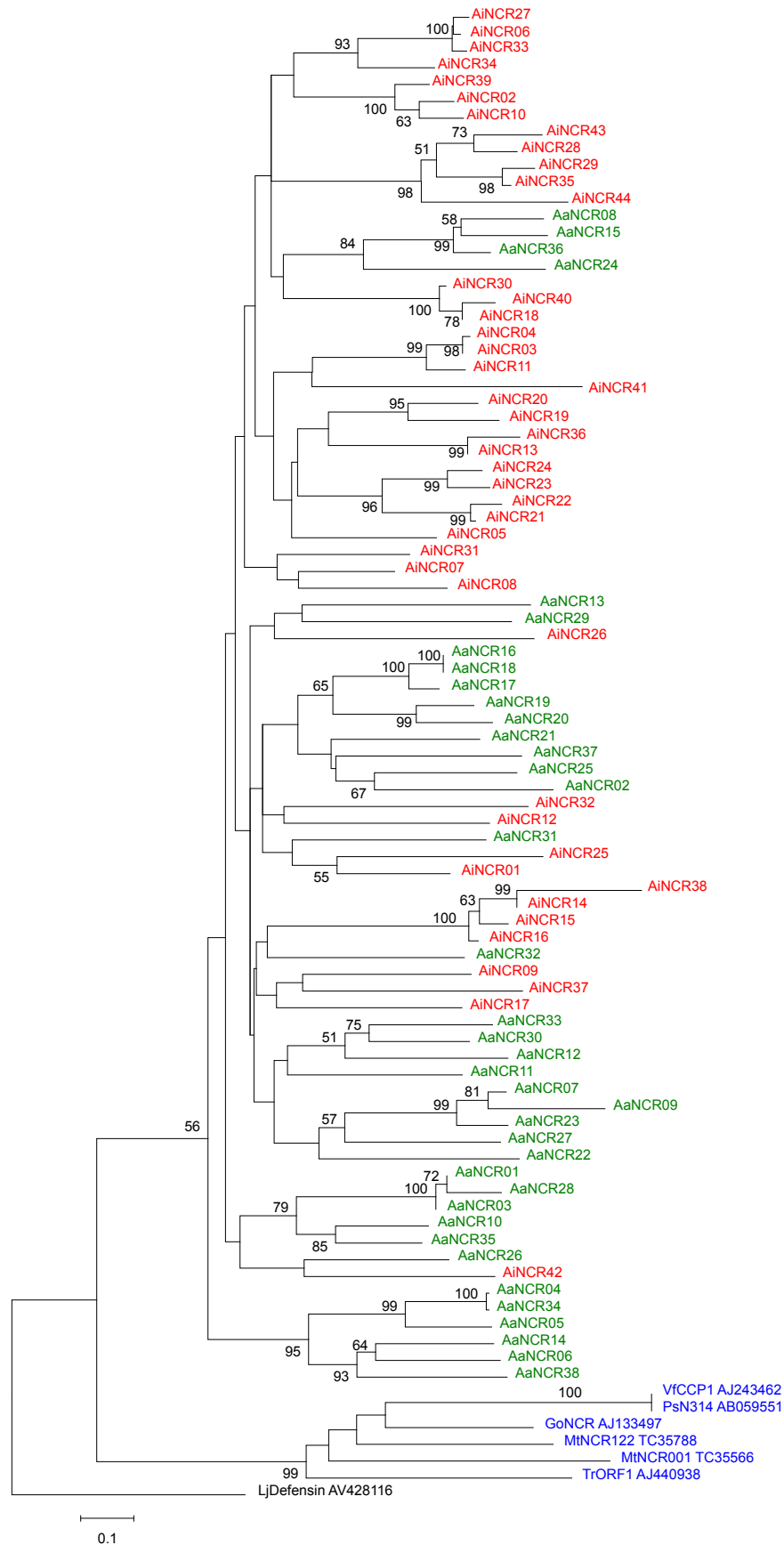
C



Supplemental Fig. S5. *In vitro* effect of synthetic *Aeschynomene* and *Medicago* NCR peptides.

(A) Effect on the viability of BTai1, hopanoid BTai1 mutant (BTai1Δshc) and ORS285 strains. The viability test was done according to Van de Velde *et al.* (2010) with 10 μM of each NCR. Control: cells treated with H₂O or NH₄OH 5% used to solubilize *Aeschynomene* anionic NCRs.

(B) and (C) Effect of AaNCR01 on *S. meliloti* bacteria. (B) Effect on the viability. The viability test was done according to Guefrachi *et al.* (2015) with 7.5 and 30 μM of AaNCR01. (C) Flow cytometry analysis of DNA content in *S. meliloti*. Blue (no treated), pink (treated with 10 μM of AaNCR01 during 4 h). Error bars in B and C are standard deviations.



Supplemental Figure S6. Neighbor joining phylogeny of NCR protein sequences. The amino acid sequences of NCR genes were aligned using Muscle3.6. The evolutionary distances were computed using the Poisson correction method, and a tree was inferred from the distance matrix by Neighbor Joining using MEGA5. All ambiguous positions were removed for each sequence pair. There were a total of 210 positions in the final dataset. Percent of bootstrap replicates (1000 replicates) are indicated at each tree node (only if >50%). NCRs from *A. afraspera* and *A. indica* are in green and red respectively and those from different IRLC legumes are in blue (The accession numbers are indicated). A defensin from *Lotus*, a non-IRCL legume was used for tree rooting.

Supplemental Table S1: *Aeschynomene afraspera* and *Aeschynomene indica* root and nodule expressed sequenced tags (ESTs) and cluster collection statistics.

	<i>A. afraspera</i>		<i>A. indica</i>	
	Root	Nodule	Root	Nodule
Number of cDNA sequenced	9492	9582	9621	9847
<u>EST summary:</u>				
Number of high-quality ESTs	9017 (95%)	8688 (91%)	9117 (95%)	8768 (89%)
Average EST length (bp)	551	513	528	510
EST size range (bp)	102-819	102-817	101-803	101-782
Non valid sequences	475 (5%)	894 (9%)	504 (5%)	1079 (11%)
Small size	360 (76%)	729 (82%)	387 (77%)	911 (84%)
Low quality	115 (24%)	165 (18%)	117 (23%)	168 (16%)
<u>Cluster summary:</u>				
Number of clusters	923	732	1006	812
Number of singletons	4162	3493	4566	4164
Number of contigs	1161	903	1268	1025
Average contig length (bp)	750	723	729	703
Contig size range (bp)	220-1794	122-2267	146-1268	120-1025

Supplemental Table S2. Nodule- specific Cys-rich putative peptides from *Aeschynomene indica*, *A. afraspera* and *A. evenia*. The NCR were named with the initials of the plant species and numbered according to the read count level in the EST nodule libraries (for *A. indica* and *A. afraspera*) or arbitrarily for *A. evenia*. The NCR with a read count above 5 in the nodules are grey shaded. Signal peptides were determined using SignalP 4.0 (Petersen TN, Brunak S, von Heijne G, Nielsen H - 2011- SignalP 4.0: discriminating signal peptides from transmembrane regions. Nature Meth 8(10):785-786) and adjusted manually. Theoretical pI and MW were calculated for mature peptides from ExPASy web site (Gasteiger E, et al. -2005- Protein Identification and Analysis Tools on the ExPASy Server; In John M. Walker (ed): The Proteomics Protocols Handbook, Humana Press). For the *A. evenia* NCRs, the pI and MW were calculated for the mature peptide even if the signal peptide was truncated or contained stop codon (indicated in the table). The presence of a signature corresponding to NCR motif 1 (1) or motif 2 (2), as defined in figure 3, is indicated in the last column. The 1* indicate a motif 1 where the spacing between Cyst 4 and 5 contains 5 amino acids instead of 4.

Sequence		Mature peptide (aa)	Nbr of Cys	pI	Theoretical MW	Read count Root	Nodule	NCR motif	
A. afraspera									
AaNCR01	CL3contig5	MAKLNWANLFAIILLVTVMSMASHLHVGLRMIDSWIDPKCLDEMCSMSSCRNKHGGLAMGHCVGSICICPFDK	46	6	6.87	5087.04	1	242	1
AaNCR02	CL13contig1	MAKLANFFTLVLLVAVISITSHHQVAESRLCNHRWFIEIRCNVNESFCNDKCSGRHGDRNTINAFVIGYCSYFDCPDGEDCSCIGF	59	10	5.44	6768.5	0	71	2
AaNCR03	CL3Contig3	MAKLNWANLFAIILLVTVMSMASHHHQVGLRMIDSWIDPKCLEMCSMSSCRNKHGGLAMGHCVGSICICPFDK	44	6	7.79	4874.83	0	45	1
AaNCR04	CL16contig1	MARLTLINLLSLVLLITGILITSSDYTVAGYVICVWRDPNCGKGDPNYCAQDKCKNDYGEKAIPVCRYDVCFCGYECTFPNNGDDYN GERNQNNRM	66	8	4.57	7615.40	0	25	2
AaNCR05	CL16contig2	MAKLTINLLSLVLLVTGMLMTSSDYKVAAGYMCVVGWRDPNCGKGDPNYCAQDKCKNDHGEKATAFCRDDVCFVLWL	46	6	4.89	5202.85	0	23	no
AaNCR06	CL31contig1	MARLTLINFLPLLLVTVGMWVKSHHYKAVGEYIDVCFDWDATNCEKGDNSFCVQCKEKNKHIEGAAQCRENRCFIFCAPNIPS PISYDGGDDYNGEGINNDGI	77	8	4.08	8573.32	0	22	2
AaNCR07	CL50contig1	MTKLTSTNIFALILLFTAISMTSLHQVANSNCWHSWMDTNPCEPETCVQLCSDSYKVLIDAFENLHVCVEYECPDN	47	8	4.01	5479.0	0	14	2
AaNCR08	CL64contig1	MTKLIILGNFFAIVLLMTVILIIKSHIEVAGEAGCEQSSDYLPIADYCEEGFDHWYDRNCPVTDFCKLECIKRGHRTGDGACNWIY CYCRYLPNSTSEDALRP	75	8	4.60	8790.69	0	10	1
AaNCR09	CL67contig1	MTKLTSTNIFALILLFTAISMTSLHQVANSNCWHSWMDTNPCEPETCVQLCSDSYKVEASRWRPMLLVKDEAPATVRDSLLQESVTV APNLWDFKIKRGPESGIN	75	4	4.94	8570.6	0	10	no
AaNCR10	CL3Contig2	MAKLNWAKFVALILLVTAMSMTHHHQVVGEMKDTWIDLRCFTDLPCTSDCRNKYGVFAVGRVGVLCCTEYELC	47	7	4.46	5372.21	0	7	1
AaNCR11	CL132contig1	MARLTVVYLFSLLLVTAISITSHHQVVGFLCLDDVWDESACERGSDGCMQDKCNKHGYTAQGFQYDSCFCSYQCS	49	8	4.08	5656.15	0	6	2
AaNCR12	CL746contig1	MSRLTSAYFFSIVLLVEAISKSHHQVAVANKLDDWFDHNKYCVEDDCVQCECKRQHNSTEGYCAYNCFPCYCEPDSPPMPPNSP VSPDYDDSPNPPTWDK	72	8	4.14	8448.1	0	6	2
AaNCR13	CL127Contig1	MTKLIILATIFALILLMIAISMTSEVDGNMCRFRWYDLNCKDPEYRGCMMCECKSNYGDYAYGFCQYKPCSNYCGPD	50	8	7.67	5959.77	0	6	2
AaNCR14	CL111Contig1	MARLTLNLFALVLLVTVGMWMTSEYKAVAGGIYTCSEVWDDPNCEKEDYTNCTQNCKRKYGMKATSHCRGDKCFICEFDLPLFS PINYDGDYGNNEGDDYNSIGINNNNDGT	85	8	4.22	9642.35	0	6	2
AaNCR15	CL211contig1	MAKLIILGNFFAIVLLMTVILIIKSHIEVAGEAGCKPSSDYIAGFMFWMNCPVTDFCNSECKRKHKPGDGECSGIYCHCRYLESEI RRDEMMS	66	7	5.19	7654.68	0	5	1
AaNCR16	CL220Contig1	MAKLSNLFALVLLVTAISMPSHHQVESKKTDEWYDIRCNLNEGYCRVDCLEVHGGGLAKGYCQGDYCFDYYCPGLLIV	52	8	4.70	6009.86	0	4	2
AaNCR17	CL263Contig1	MAKLSNFFALVLLVTAIWMPSHHQVEGKECTDRWYDERCSLNEGYCRVDCLEVHGGGLAKGYCQGDYCFDYYCPGLLIV	53	8	4.75	6219.02	0	4	1*
AaNCR18	CL220Contig2	MAKLSNLFALVLLVTAIPMPSHHQVESKKTDE*YDIRCNLNEGYCRVDCLEVHGGGLAKGYCQGDYCFDYYCPGLLIV	51	8	ND	ND	0	4	2
AaNCR19	CL266Contig1	MAKLANFFTLVLLVTAISMTMNQVVEGKICNEMWYDIRCNLSESDCRLDICIELYGGGPAKGYCVGDYCCCNVDCVGDYCSNYE QEQVCGGPQ	66	12	4.02	7420.27	0	4	2
AaNCR20	CL136Contig1	MAKLANFLTLVLLVTAISMTMNQVVEGKICNEMWYDIRCNLSESYCREDCIEKHGVLAKGYRGMGNKCCCLTV	45	7	7.73	5279.24	0	4	no
AaNCR21	CL297Contig1	MAKLANFFTLVLLVTAISMTSHLQVEGWCKNKWLDIRCGLYESYCWEDCSKLHGETAIGFCEDIYCSYDSCRNIKN	52	8	5.09	6258.06	0	4	2
AaNCR22	CL181Contig1	MTELTSKNHFALIIIVTAISMRSLLHQVARECIDTWIDRNCEYNYIEICKLDCKDYGGESSFAYCVDHHKCKCHYKCPQWPYFN	55	8	5.82	6841.71	0	4	1*
AaNCR23	CL50Contig2	MTKLTSTNIFALILLFTAISMTSLHQVANSNCWYSWMDKYCEDETQVQSCNDTYKVLIDAFENIHCFCYECPNKDK	49	8	4.28	5918.59	0	4	2
AaNCR24	CL217Contig1	MTKLTLANIFAIVILVIMLIQSHDGVAGQDDYCEPNSDYLPQIANCWEDGLYSWFDLTCPMSNENCKLDCKRKVGTPTGDGVCL IYCYCRY	64	8	4.20	7493.43	0	4	1
AaNCR25	CL226Contig1	MAKLAKFLPIVLLVTAISITSHHQVAGKLCNHRWFDIRCNLVDYDNESFCNKDCSLHGDVAMFGFVLDYCSCHYVSPVTV DEHCSNYEQNGLCVGV	77	10	5.03	8817.89	0	4	1
AaNCR26	CL779Contig1	MVKLTLTNLFAVLLVTVGMLMRSDDVGEVTFDWDWNCRVDEINCMIDCQEKHGITAVGSCIAIICKVCFESNC	48	7	4.28	5455.30	0	3	1
AaNCR27	CL416Contig1	MVKLTLTNFALVLLVTAISMGSHQVQASDLWNSWIIINCKEDCKKQCSSEYVPIIEAFAGSYCFCKYEET	45	8	5.06	5267.11	0	3	2
AaNCR28	CL3Contig1	MSMASRHHQVGLRMIDSWIDPKCLDILGLIQSAYEMSCMSSCRNKHGGLAMGHCVGSICICPFDK	55	6	6.87	6046.19	0	2	1
AaNCR29	CL491Contig1	MTKLTLANFVLSLQITVIMWTSNGKAVAGEDCLCNWFDKNCRDGDSKECVKDCRDSHGNNMAHGSCAYDLVCSYSCPNSSPS WDK	56	9	4.74	6286.88	0	2	2
AaNCR30	CL820Contig1	MTNLDLTNFFAILLVTASMTSHHHQVVEVCLDEFDMNCREIDSTICFEDCKYMHGKTKGVCAFNHCTCMYEC	46	8	4.82	5395.21	0	2	2
AaNCR31	CL414Contig1	MAKLNKTFFAVLLIIAISMTSHHQVAGDVCYEWADFLCLFGYSRCSRYCEFAHQGVGATIGYCEDGKCTCSFQC	49	8	4.58	5587.25	0	2	2
AaNCR32	CL371Contig1	MAKLTNTNFVVLVSVISMTSDHHQVVGSKCPDEGVDNKYCSSEWVDMNCTTGSEDCSQECVNNYIGAEAGSCAGTKCCCRY YCPGSLE	62	10	4.05	6723.35	0	2	2
AaNCR33	CL310Contig1	MNTLTLANFFAVILSVTAISMTSYHHHQVVEECLDMFMNCEIEDSNKNVCEACKYMHAGDNNVGDGKVLVDHCYCKYEC	49	8	4.27	5676.35	0	2	2

Sequence		Mature peptide (aa)	Nbr of Cys	pI	Theoretical MW	Read count Root	Nodule	NCR motif	
AaNCR34	CL16Contig4	MARLTLINLLSLLITGILITSSDYTVAGY YVIVVWRDPNCGKGDPNYCAQDCCKNGYGEKAIPVCRYDVCFCGYECTFPNNGDDYN GERNQNNRM	66	8	4.77	7557.36	0	2	2
AaNCR35	CL3Contig4	MAKLWAKFVALILLVTAAMLMTSYHHQVAGI HTTDSWIDPSCFDEMFCMLDCRNKHGVFALGSCVGNICECTYELKS	48	6	4.63	5422.18	0	1	1
AaNCR36	CL304Contig1	MTKLILGNFFAIVLLMTVILIIKSHIEVAGE ASCEPSSDYLPQIAGFDSWYDINCPMSDFLCNMECKRKYGRGTGYGGCVGNCHCY	58	7	4.65	6676.44	0	1	1
AaNCR37	CL278Contig1	MAKLATFLHFFWSNIFALFLVTAISITHHQVAC GKLCNHHKWFDRRCGYMDDSDCNKYCLDLHGDKLAGFCVYVYCSCHYICPG DDVPVDEDCSCNYEQKTCVCGWSLGSCTNIVLPLVVG	89	13	4.67	10038.41	0	1	2
AaNCR38	CL77Contig1	MARLTLINFLPLLLVTGMWMTSQYKVAGS YICVYVWDDPNCKNEDYTNCTQECESKHGKEARPICREEKCFRCDFDFFPLPTPN DDGNYGNNDGI	68	8	4.37	7972.70	0	1	2
A. indica									
AiNCR01	CL27contig1	MTKLTLLNFFTFGLLITAMWMTSHHQVFG EQCIDMWFDGCRNLDLHFKCSEDCGKHAIGECADEYCFCSYECDNNSPK	51	8	4.41	5875.44	0	21	2
AiNCR02	CL19contig1	MIKLTLLNFALVLLITAMSMINHGEAA GTGFDSWVDKNCQENFYCIHSCRQKHGQSAMGEGGGSFCYCY	46	6	6.02	5242.79	0	19	1
AiNCR03	CL25contig1	MTKFLTNVFAVLLVTVISITSHHQVGTI AFDSWVDFYCATDNYNCIMSCRNKYGITGRGNCNDVFCVCFYKLLN	48	6	6.06	5563.34	0	17	1
AiNCR04	CL25contig2	MTKFLTNVFAVLLVTVISITSHHQVGTI AFDSWVDFYCATDNYNCIMSCRNKYGITGRGNCNDVFCVCFYKLLN	48	6	6.06	5563.34	0	8	1
AiNCR05	CL84contig1	MTKLTLLNFFTLISVTAIITSHHQVASM GFDTWIDIDCPLNNHSCIMNCRKYGLTASGKNCNDVFCFFYLLN	47	7	5.34	4250.84	0	8	1
AiNCR06	CL45contig1	MIKLSLINFALVLLITAILMISHHEVA CLGSDSWMDINCPGDDSDICNFHCKRKYGSSAWAMCSGSFCRCYH	43	6	6.87	4840.44	0	7	1
AiNCR07	CL164contig1	MIKLTLLNFALVLLVAVISSHHQVDSI QRDYVWDINCYNCVMDCKINHGDTAVGRCVGNLCYCTY	41	6	5.36	4799.50	0	7	1
AiNCR08	CL343contig1	MINLIFTKFFALVLLMTAISHHQVGTI EYDYVWDMCNVCAIDCVKHHGVTALGKVCNLCYCRYNMS	44	7	6.70	5119.02	0	7	1
AiNCR09	CL107contig1	MVELTLVNFALVLLITDISMTHHHQVA AFYFCLNDVWVDENCKEKGVSINCMEDCINKYGDTSNGFCFYNCYCQYKCPSPDSP NSKDSPIEPHN	67	8	4.13	7838.57	0	7	2
AiNCR10	CL19contig2	MTKLTLLNFFTFVLLMTAISMINHGKA AAGTGFDSWIDKNCPMENFYCIHSCRQKHGQSAMGEGGGSFCYCY	46	6	6.02	5289.90	0	6	1
AiNCR11	CL150contig1	MTKLTLLNFALVLLMTVISIISHHQVGTI AFDSWVDFYCATDNYNCIMSCRNKHGITGMGNCKDVFCVYCYLLN	48	6	5.36	5528.31	0	6	1
AiNCR12	CL46contig2	MTKLTLLKFFAILLVTAAMSMINHQA AGATCEIETDLRCKLYDGGCWMDCKLYGSDASRGHCQDDYNYFCLEYCRPKYLAK	54	8	6.74	6367.34	0	6	1*
AiNCR13	CL176contig1	MIRLILANFIVLVLVITISIRKHKVQV SGFECWIDLDCPWENHTCIVSCRNKHGLTASGNCYDMFCYCYTY	45	7	5.35	5355.06	0	5	1
AiNCR14	CL698contig1	MVKLTLLNFALVLLMAVTSKTIHHQVIG YDCVEDWIDENCKEKGDYSTCFDFCDVKHGDIAIGYCEGHNFQYGSCTHIKDHIDIGY	59	8	4.25	6795.41	0	5	2
AiNCR15	CL425contig1	MVKLILNFALVLLMPPVTSKTIHRQVA AYDVCENWIDENCKEKGDYSTCFDFCDVKHGDIAIGYCEGHKFCQYGSCTHTKDHID IGY	59	8	4.49	6814.48	0	5	2
AiNCR16	CL88contig1	MVKLILNFALVLLMSVISKTIHYQV TYDCVENWIDENCKEKGDYSTCFDFCDVKHGDIAIGYCEGHKFCQYGSCTHTKDHID GY	59	8	4.49	6814.48	0	5	2
AiNCR17	CL172Contig1	MTKVLTLNFALVFLITVISMTRHHQV AGYLGLDDWIDENCTRLFFSDSGCVKDKCKHKGDTGDGYCVDNRCFSYIEPDSPNYGD SPDWSE	63	6	4.15	7192.72	0	5	1
AiNCR18	CL286contig1	MSKLTLANFFSIVLLVTVILMITRHQV TSEEILDSWIDINCPMDNFNCNWDCVHKHGATANGYCNNGNICCKYFLHS	48	7	4.83	5513.18	0	5	1
AiNCR19	CL615Contig1	MIRLILANFFTPVLLMSAISITGKHQV ATLGFATWIDGHCPRENLCIVNCRSEYGPSAKGMCDNVFCFYTYI	46	6	5.48	5276.03	0	4	1
AiNCR20	CL848Contig1	MTRLTLINFFIPVLLMSAISTTSQHQV AGGVGFANWIDLECPANFYCMIDCRSEYGPSAKGLCNDVFCFYTYIN	47	6	4.44	5371.13	0	4	1
AiNCR21	CL521Contig1	MIKLTLANFFSILLVAAISMSTSQHQA VAFSGFDKWDIDFDCPVTSLNCIADCRERYGALTRAYCINNVCFCFYTYQY	46	6	5.47	5487.19	0	4	1
AiNCR22	CL907Contig1	MTRLTLANFLIILLVAAISMSTSQHQA VAFSGYDKWIDFDCPVTSLNCIADCRERYGALTRAYCINNVCFCFYTYQY	48	6	5.47	5737.44	0	4	1
AiNCR23	CL288Contig1	MTKLILANFFILVLLMTAISMTSQYQV ADSGYDKWINLDCEVTLLNCIADCKRKYGPARIAYCSNIYCYFYKY	47	6	8.22	5583.40	0	4	1
AiNCR24	CL605Contig1	MTRLILANFFTLVLLMTAISMTSQHEVA DSGYDKWIDLYCEVTSLNCAIECKRYGPIARAYCSSNIYCYFYSKY	47	6	8.22	5556.38	0	4	1*
AiNCR25	CL255Contig1	NLHVLYHFNSYLAAIWMAAGDHQVY GDVCFDHFDFLGGCKNGDFDRCLDCTKHGNLATGEDEYDICYRYECCDKPYPLS	55	9	4.26	6391.05	0	4	2
AiNCR26	CL269Contig1	MIKLTLANFIVLVAISMKSNC EVATTEDLARWYDSNLESDKTQIECKFRYGITAQGGCCDFLCCYFLMPDLVK	53	8	4.46	6071.00	0	4	1*
AiNCR27	CL45Contig2	MIKLSLTLNFALVLLITAILMISHHEVA CLVSDSWMDINCPGDDSDICNFHCKRKYGSSAWAMCSGSFCRCYH	43	6	6.87	4882.52	0	3	1
AiNCR28	CL92Contig2	MTNKLISLANFVAFMLLVTAISSSHNP AMGILPCFDYWTDKNCPDNDNDCIQDCKDKYGNVVGNCNGYCYCRYACPRYTRME GPHVLVDKVNINIDGMQYR	75	8	5.59	8716.81	0	3	2
AiNCR29	CL339Contig1	MASKLTLANFVFMILLTAIWTSSHNP AMGSI PCYDYWTDPNCPDDDDYCIQVCKERYGNSAVGNCNGCYCRYSCPRYT	51	9	4.27	5957.55	0	3	2
AiNCR30	CL396Contig1	MSFLLTNFFSIVLLVTAISMTHHQV TSEEILDSWIDINCPDNFNCNWDCVKNHGATANGYCNNGNICCKYFLHS	48	7	4.51	5472.10	0	3	1
AiNCR31	CL295Contig1	MTKLTLLNFALVLLVTAISITSHHHVNS IQCDFDNDWIDICINNCSMDCIEKYGTALGKCVGKTYCCYESMS	45	8	4.23	5118.90	0	3	1
AiNCR32	CL468Contig1	MNKLINWACIFGVFVVTAIMTNEVA GEDCLDFWNPDRCKYSNQYICKTNCKTKHGDLAGSCLDRVCYAYVC	47	8	6.77	5460.24	0	3	2
AiNCR33	CL45Contig3	MIKLSLTLNFALVLLITAILMISHHEVA CLGSDSWMDINCPGDDSDICNFHCKRKYGSSAWAMCSGSFCRCYH	43	6	6.87	4874.46	0	2	1
AiNCR34	CL831Contig1	MIKLTLLNFALVLLVTAIILISHREVA SFGSDSWIDINCPDNDNFCNCRKYGSSAWAMCSGNFCHCY	43	6	7.78	5082.65	0	2	1
AiNCR35	CL909Contig1	MASKLTLANFVFMILLMTAISSSHNP AMGSI PCYDYWTDPNCPDDDDYCVQVQERYGNSAVGNCNGCYCRYSCPRYT	51	9	4.04	5943.48	0	2	2
AiNCR36	CL905Contig1	MTRIDFANFFTLILLIHWIIHQV GTGFECWIDLDCPWENHTCIVSCRNKHGLTASGNCYDMFCYCYTY	46	7	5.33	5456.16	0	2	1
AiNCR37	CL941Contig1	MSRLTLINFAVILLVIVSRNHLVAGD LCYEEVWDENCINNGSYCMRDKCKKYGNTGDGLCALLKCVLYECPDSRSDSPEPRI	58	8	4.50	6693.55	0	2	2

Sequence		Mature peptide (aa)	Nbr of Cys	pI	Theoretical MW	Read count Root	Nodule	NCR motif
AeNCR42	CL9423Contig1	KLNLVFN*YFAVISMYSYHHVDSIQCFDNWVDIDCINNCMSDCIEKYGFTVLGKCVGKTCYCCYE	42	8	4.23	4841.60	stop codon	1
AeNCR43	CL4722Contig2	MTKLTLLINFFMFLVLLVTAMLMTSYDQVSGEVCFRVWYDLSCYRGNNSHRCSMACTNKYGDLAGMGRCDRGHCNCRYESEEC	50	8	6.93	5879.59		1
AeNCR44	CL14243Contig1	MTKLTLAYIFVLLVLTVAISMTNEVAGEKKEEFWYEPQCKCPNYNGCMTDCKIKHGDSAKGACDYLDYCYTYD	46	8	4.56	5373.00		1
AeNCR45	CL26264Contig1	MTKLTLAYIFAVLLVTAISLITEVAGEECLDYWYDRRCRRRNYKGCMRDCKDYGDSANGFCDFLDFCFCSFD	46	7	5.12	5670.31		1
AeNCR46	CL11729Contig1	MTKLTLLINFFTFALLVAGILMYSYHQVSGEVCFFHLWSDLNCHKGDSLHCLVACTNKYGDLAGKRCSDYCNCRYESEK	50	8	5.80	5734.43		1
AeNCR47	CL6739Contig2	MIKLTLANIFLLTLLMTVISRTSSCEAAHIDGFVRWDFSCAENDKSCHECLRLRFGATAQGFCEDEVFCCCN	46	7	4.96	5315.97		1
AeNCR48	CL3731Contig2	YFAVIMINRVEMAGTGFDSWVDKNCPLNFYCIHSCRQKHGQSAMGECGGSFYCY	58	6	6.02	5064.64	signal peptide truncated	1
AeNCR49	CL13080Contig2	MAKLTSTNFFALLTLLISVILMISHHEATEMGFESWIDIDCPVDFHCKDICGKKGPTTSAEKNEIFCYCY	45	6	4.28	5213.86		1
AeNCR50	CL682Contig4	MNKLNWACIFGLVFFVTAIIMTNEVVGEDCLDFWSDPRCKYSNQYICKTCKKHGDLAIGSCLDRVCYCA	45	7	6.78	5230.94		1
AeNCR51	CL12848Contig1	FTVILITSHHQVDSIQRYDYWVDINCIYNCMIDCINKHGDTALGR	34	4	5.36	3966.55	signal peptide truncated	no
AeNCR52	CL2051Contig2	MSKLTLAIFPLVLLVTVMMVMSIPCFDYWSNMNCPDDDDNCIQCTCKDRYGD	42	7	3.69	4828.20		no
AeNCR53	CL4958Contig2	IFYFN*YFAAIWIKNHGQVAGDLCFEEWVDENCINNGSYDCMRDCKNKYGTGDLCAVLKCVCLYE	43	6	6.70	5043.62	stop codon	1
AeNCR54	CL682Contig1	MNKLNWACIFGLVFFVTAISMTSEVAGEDCLDVWYEPHCFEQKSNVCKKNCKSKHGLANGFCNASDLICY	44	6	5.44	5056.68		no
AeNCR55	CL1Contig16	LFYIT**KDAISTSSHNPAMGSI	46	8	3.80	5336.78	stop codon	1
AeNCR56	CL10596Contig3	MQQLTANFFALLVLLGIVISMISHHVAGDL	26	4	3.61	3107.42		no
AeNCR57	CL5726Contig1	LSFN*YFAAIWIKNHGQVAGDLCFEEWVDENCINNGSYDCMRDCKNKYGTGDLCAVLKCVCLYE	46	7	4.16	5247.90	stop codon	1
AeNCR58	CL4882Contig4	FAAISMISHHQVAGVTCIEIWDLRCKLYDGGCWLDC	48	8	5.57	5590.43	signal peptide truncated	1*
AeNCR59	CL2051Contig4	LINFAAILMSSHNSVMVMGSI	45	8	3.90	5250.70	signal peptide truncated	1
AeNCR60	CL2051Contig3	TILMSSHNSVMVMGSI	42	7	3.69	4828.20	signal peptide truncated	no
AeNCR61	CL27458Contig1	IFALVFLVTAISMTSNRYVTSTRFSKSWIDVQCPKEDLNCSNLCNENYGP	61	5	4.86	4644.32	signal peptide truncated	no
AeNCR62	CL682Contig3	FLALIFLLTAISMTSEVAGEDCLDVWYEPHCFEQKSNVCKKNCKSKHGLANGFCNASDLICY	44	6	5.44	5056.68	signal peptide truncated	no
AeNCR63	CL11599Contig2	NMHVIFHLN*YFAVTSMTNHHQIAGIMCTKYWFD	33	4	8.49	3940.56	stop codon	no
AeNCR64	CL12457Contig1	MVKSTLATFFVILLVTAIIMTSHHPVMAGYCV	48	7	4.24	5657.32		1
AeNCR65	CL7951Contig1	IFALVFLVTAISMTSEATGEDCLHIWYEPQCEFRNHNVCNMNCKTKHGDI	65	7	4.72	6237.01	signal peptide truncated	1
AeNCR66	CL24618Contig1	MTKLTLLINIFALVFLVSVISMISDVAGKKGQSWYDLQCKDPNYNGCTKDCINKY	29	4	8.45	3373.80		no
AeNCR67	CL19676Contig1	MIKLTLANVFLVFLVSVISMISDVAGKKGQSWYDLQCKDPNYNGCTKDCINKY	46	7	6.86	5708.56		1
AeNCR68	CL3827Contig1	ILHFY*YFTAILMMISHHEVA	43	6	6.87	4868.45	stop codon	1
AeNCR69	CL10339Contig1	MTKLTLANNFALLVAVISMISDVAGDGKIGKVVYDAQCQY	47	7	4.86	5615.29		1
AeNCR70	CL4722Contig1	LTKIQTCNL*FEFIVAMLMTSYDQVSGEVCFRVWYDLSCYRGNNSHRCSMACTNKYGDLAGMGRCDRGHCNCRYESEEC	50	8	6.93	5879.59	stop codon	1
AeNCR71	CL682Contig2	FVAIIMTNEVVGEDCLDFWSDPRCKYSNQYICKTCKKHGDLAIGSCLDRVCYCA	57	7	6.78	5230.94	signal peptide truncated	1
AeNCR72	CL8283Contig3	DTWIDQNCPWESFYCRLSCRTHKGGKAMGECGGGPFYCY	41	6	ND		ND signal peptide absent	1*
AeNCR73	CL24495Contig1	LSFN*YFVISTKNHHLVAGDVCFEWVDLNCINNDPNYCMRDCKNKYGD	36	5	3.98	4163.58	stop codon	no
AeNCR74	CL1101Contig4	MTKLIVANIFSVLLVTAIIMRSEVVDASSELVYDINCKDPGYNTECMQYCIDKHGYTARGSCLYNNYCYC	44	6	4.83	5073.64		1*
AeNCR75	CL1159Contig4	MIKLTLLINFFALVLLITDDWFDENCEKISIDCMEDCINKYGDTSNGFC	59	6	3.97	4895.36		1
AeNCR76	CL6739Contig3	FAVISRTSSCEAAHIDGFVRWDFSCAENDKSCHECLRLRFGATAQGFCEDEVFCCCN	58	8	4.95	4822.45	signal peptide truncated	1
AeNCR77	CL6739Contig1	LKYHVISRTSSCEAAHIDGFVRWDFSCAENDKSCHECLRLRFGATAQGFCEDEVFCCCN	60	8	4.96	4822.46	signal peptide truncated	1
AeNCR78	CL13080Contig1	YFAVILMISHHEATEMGFESWIDIDCPVDFHCKDICGKKGPTTSAEKNEIFCYCY	58	6	4.38	4795.39	signal peptide truncated	1
AeNCR79	CL15279Contig1	IYKFLNI*FVAISMYSKVDGKECLAVWYEPQCEYPNNVCMMNCKTNHGYLAGVCGDIFCYCNYD	67	7	4.50	5277.99	stop codon	1
AeNCR80	CL25922Contig1	FVAISLITEVAGEECLDYWYDRRCRRRNYKGCMRDCKDKYGD	58	7	5.12	5670.31	signal peptide truncated	1
AeNCR81	CL11729Contig2	KIQTHNLSF*FIFAGILMYSYHQVSGEVCFFHLWSDLNCHKGDSLHCLVACTNKYGDLAGK	34	4	6.27	3778.31	stop codon	no
AeNCR82	CL1101Contig2	TLFLFLIEDAAILMRSEVVDASSELVYDINCKDPGYNTECMQYCIDKHGYTARGSCLYNNYCYC	65	6	4.83	5073.64	signal peptide truncated	1*

Gene name	Protein function	Protein mass (Da)	ORS285 free-living state	Bacteroid A. indica ORS285	Bacteroid A. afraspera ORS285	ORS285 free-living state				Bacteroid A. indica - ORS285				Bacteroid A. afraspera - ORS285			
						Mascot Protein score	Protein coverage (%)	Peptide sequences	Peptide modifications	Mascot Protein score	Protein coverage (%)	Peptide sequences	Peptide modifications	Mascot Protein score	Protein coverage (%)	Peptide sequences	Peptide modifications
BRAO285v1_1200089	conserved protein of unknown function	7603	125 / 31	678 / 98,6 229 / 50,7	198 / 45,1 602 / 98,6	125	31	MDNITLHVR FAPDGTVAEISER	Oxidation (M)	678 229	98,6 50,7	AYQTFAGGR MDNITLHVR DQVEALKATAAA FAPDGTVAEISER AYQTFAGGRGVFK TTRDQVEALKATAAA LSEAIGLKAYQTFAGGR PATLTPQQWFNKLSEAIGLK FAPDGTVAEISERPATLTPQQWFNK GVFKTTR MDNITLHVR TTRDQVEALK FAPDGTVAEISER	Deamidated (NQ) Oxidation (M) Oxidation (M)	198 602	45,1 98,6	MDNITLHVR TTRDQVEALK FAPDGTVAEISER AYQTFAGGRGVFK PATLTPQQWFNK TTRDQVEALKATAAA LSEAIGLKAYQTFAGGR PATLTPQQWFNKLSEAIGLK FAPDGTVAEISERPATLTPQQWFNK	Oxidation (M) Deamidated (NQ) Oxidation (M)
BRAO285v1_1200091	conserved protein of unknown function	5597	55 / 17,9	76 / 17,9 108 / 64,3	105 / 64,3	55	17,9	LIEALASMPK	Oxidation (M)	76 108	17,9 64,3	LIEALASMPK VDALVELVAEQGPGAPAEVILQPSLDE	Oxidation (M) Oxidation (M)	105	64,3	LIEALASMPK VDALVELVAEQGPGAPAEVILQPSLDE	Oxidation (M)
BRAO285v1_1200098	conserved protein of unknown function	8237	/	262 / 67,9	64 / 11,5 342 / 83,3					262	67,9	AMAATDIER GQMGGLVHALALETSLPK EAFPPDAVVVIEDLAGDGDHYSAR LIKEAFPPDAVVVIEDLAGDGDHYSAR	Oxidation (M)	64 342	11,5 83,3	AMAATDIER GQMGGLVHALALETSLPK EAFPPDAVVVIEDLAGDGDHYSAR VQQHQVMVYSALGQMGGLVHALALETSLPK	Oxidation (M)
BRAO285v1_1200103	nifH nitrogenase iron protein, nifH	31894	/	114 / 10,5	/					114	10,5	CVESGGPEPGVGCAGR VLEYPADSVQAGHYR					
BRAO285v1_1200113	mop putative Molybdenum-pterin binding protein (Mop)	6882	63 / 17,6	235 / 79,4 63 / 17,6	51 / 17,6 155 / 61,6	63		NVLPKVISVAK		235 63	79,4 17,6	NVLPKVISVAK VELAPGLTVFSAITNEAVEELGLAVGDQVSAVIK GATTAHVKVELAPGLTVFSAITNEAVEELGLAVGDQVSAVIK NVLPKVISVAK		51 155	17,6 61,6	NVLPKVISVAK VELAPGLTVFSAITNEAVEELGLAVGDQVSAVIK GATTAHVKVELAPGLTVFSAITNEAVEELGLAVGDQVSAVIK	
BRAO285v1_1200160	conserved protein of unknown function	6926	72 / 20,6	78 / 20,6	79 / 20,6	72	20,6	LGAPSEAVESLR		78	20,6	LGAPSEAVESLR		79	20,6	LGAPSEAVESLR	
BRAO285v1_1220114	putative periplasmic mannitol-binding protein	40397	83 / 3,9	/	/	83	3,9	SLDTIYGGAEYMAK									
BRAO285v1_1220119	conserved exported protein of unknown function	9730	50 / 13,2	/	/	50	13,2	DEYVVQGPSHR									
BRAO285v1_130041	phbA acetyl-CoA acetyltransferase with thiolase domain (Acetoacetyl-CoA thiolase)	40812	58 / 2,8	/	/	58	2,8	SDDVVIVSAAR									
BRAO285v1_1320032	conserved protein of unknown function	8776	76 / 11,1	/	/	76	11,1	VISPGSVYR									
BRAO285v1_1440053	conserved protein of unknown function	8501	/	99 / 22,5	99 / 22,5					99	22,5	SAADPGGVLCDIDTVAR		99	22,5	SAADPGGVLCDIDTVAR	
BRAO285v1_1450020	putative glutaminase	68031	49 / 1,1	/	/	49	1,1	NLAQLLA	Deamidated (NQ)								
BRAO285v1_1460012	conserved exported protein of unknown function	10997	96 / 20	/	78 / 20	96	20	GAAIGAVIGAGTGAAIAADGER						78	20	GAAIGAVIGAGTGAAIAADGER	
BRAO285v1_1520084	glnA Glutamine synthetase I (Glutamate--ammonia ligase I) (GSI)	52612	/	52 / 2,8	/					52	2,8	AIKDNVDKVVDLR					
BRAO285v1_1580051	putative endoribonuclease L-PSP (protein synthesis inhibitor) (modular protein)	15831	104 / 12,2	/	/	104	12,2	AGTVEQTLAAQINLPAPK									
BRAO285v1_1640009	putative 50S ribosomal protein L36	4987	/	68 / 21,4	/					68	21,4	VYVINKVQR					
BRAO285v1_1740028	rpsP 30S ribosomal protein S16	12407	/	55 / 9	/					55	9	FLDTAGVKKR					
BRAO285v1_190026	conserved protein of unknown function	20668	52 / 4,2	54 / 4,2 58 / 4,2	63 / 4,2 53 / 4,2	52	4,2	IASLISPR		54 58	4,2	IASLISPR		63 53	4,2 4,2	IASLISPR IASLISPR	
BRAO285v1_1910004	conserved protein of unknown function	7521	51 / 24,6	496 / 81,2	374 / 81,2	51	24,6	QITEQAASQSGAADEER		496	81,2	LAQOSELDALLK SLTDPTLSAADLADR QITEQAASQSGAADEER LAQOSELDALLKEQER LAQOSELDALLKEQER SLTDPTLSAADLADRIAILR VFDELGADKQAEAAALK	Deamidated (NQ)	374	81,2	SLTDPTLSAADLADR QITEQAASQSGAADEER LAQOSELDALLKEQER SLTDPTLSAADLADRIAILR	
BRAO285v1_1910012	conserved protein of unknown function	6797	91 / 24,2	/	/	91	24,2	EAFESAAAAATLALK									
BRAO285v1_1950033	50S ribosomal protein L29	7896	/	53 / 23,2	/					53	23,2	AMSPDQMDAIVNKK	2 Oxidation (M)				
BRAO285v1_1950040	rplF 50S ribosomal subunit protein L6	19326	/	/	61 / 5,1									61	5,1	VGVAAEIR	
BRAO285v1_1950043	rpmD 50S ribosomal subunit protein L30	7114	49 / 17,2	184 / 40,6	116 / 21,9	49	17,2	VTLPDTPAVR		184	40,6	VTLPDTPAVR TIKIQGSPR IGRVTELPDTPAVR		116	21,9	VTLPDTPAVR IGRVTELPDTPAVR	

Gene name	Protein function	Protein mass (Da)	ORS285 free-living state	Bacteroid A. indica ORS285	Bacteroid A. afraspera ORS285	ORS285 free-living state				Bacteroid A. indica - ORS285				Bacteroid A. afraspera - ORS285			
						Mascot Protein score	Protein coverage (%)	Peptide sequences	Peptide modifications	Mascot Protein score	Protein coverage (%)	Peptide sequences	Peptide modifications	Mascot Protein score	Protein coverage (%)	Peptide sequences	Peptide modifications
BRAO285v1_2000062	conserved protein of unknown function	6768	336 / 69	69 / 22,4 72 / 22,4	73 / 22,4 67 / 22,4	336	69	TNDLKLVER MTNDLKLVER SFAYGNTNIENER QIEMTPDQLGEQR SFAYGNTNIENERITR	Deamidated (NQ) Oxidation (M)	69 72	22,4 22,4	SFAYGNTNIENER SFAYGNTNIENER		73 67	22,4 22,4	SFAYGNTNIENER SFAYGNTNIENER	
BRAO285v1_200016	putative transcriptional regulatory protein	14806	/	63 / 9,1	/		63				9,1	SAGKPDIELGKR					
BRAO285v1_200029	ccoQ Transmembrane protein	6115	/	/	55 / 25,9									55	25,9	AAFDEAANMPLREE	
BRAO285v1_2010015	groL 60 kDa chaperonin (Protein Cpn60) (groEL protein)	57790	/	/	54 / 2,2									54	2,2	NVVLDSFGAPR	
BRAO285v1_2020033	cspA Cold shock protein, DNA binding	7526	/	274 / 78,9	397 / 78,9		274	SMGTVKWFNATK EGQKISYEIVADR AGLGLTREGQKISYEIVADR GYGFIQDDGGNDVVFHISAVR WFNATKGYGFIQDDGGNDVVFHISAVR	Oxidation (M)		78,9	SMGTVKWFNATK EGQKISYEIVADR AGLGLTREGQKISYEIVADR GYGFIQDDGGNDVVFHISAVR		397	78,9	SMGTVKWFNATK EGQKISYEIVADR EGQKISYEIVADR EGQKISYEIVADR AGLGLTREGQKISYEIVADR GYGFIQDDGGNDVVFHISAVR WFNATKGYGFIQDDGGNDVVFHISAVR	Oxidation (M)
BRAO285v1_2040006	conserved protein of unknown function	7257	/	334 / 80,6	121 / 16,1		334	LMVKDEIER LRQVTETLH KLMVKDEIER AIQAHLVELER VLENEHLHALVHPSTDDL	Oxidation (M)		80,6	LMVKDEIER LRQVTETLH KLMVKDEIER AIQAHLVELER VLENEHLHALVHPSTDDL		121	16,1	LMVKDEIER KLMVKDEIER	Oxidation (M)
BRAO285v1_2040025	putative transcriptional regulator/antitoxin, MazE	8451	62 / 14,7	95 / 14,7	72 / 14,7	62	14,7	IGNSDGLLLPR		95	14,7	IGNSDGLLLPR		72	14,7	IGNSDGLLLPR	
BRAO285v1_2110001	conserved protein of unknown function	9052	60 / 11,6	/	/	60	11,6	DQNSVSLLLR									
BRAO285v1_2190017	putative Isomerase/Decarboxylase related protein family		/	55 / 6	67 / 6		55	6	IAALDPASLPAVSGKPR					67	6	IAALDPASLPAVSGKPR	
BRAO285v1_2480007	cspA Cold shock protein, DNA binding		/	62 / 18,1	/		62	18,1	AMTGTVKFFNGER				Deamidated (NQ)				
BRAO285v1_2570023	putative ABC transporter (substrate-binding protein)	30727	51 / 3,1	54 / 3,1 60 / 3,1	64 / 3,1	51	3,1	VAVTSLVGP		54 60	3,1 3,1	VAVTSLVGP VAVTSLVGP		64	3,1	VAVTSLVGP	
BRAO285v1_2890040	aldehyde dehydrogenase	50682	52 / 2,7	/	/	52	2,7	AGNVNLQGVPNDR									
BRAO285v1_290022	conserved protein of unknown function		/	153 / 39,4	53 / 22,7 150 / 39,4		153	39,4	LLEILVCPMTK GPLEYDAAKQELVSR				Oxidation (M)	53 150	22,7 39,4	GPLEYDAAKQELVSR LLEILVCPMTK GPLEYDAAKQELVSR	
BRAO285v1_2940016	conserved protein of unknown function	6881	49 / 17,7	/	/	49	17,7	TAEFSEALDRK									
BRAO285v1_2940028	cspA Cold shock protein, DNA binding (modular protein)	7576	/	84 / 28,6	52 / 28,6		84	28,6	AGLTLNENQAIYDLVESR					52	28,6	AGLTLNENQAIYDLVESR	Deamidated (NQ)
BRAO285v1_2940061	conserved protein of unknown function	10103	125 / 22,2	/	/	125	22,2	VLVDDGTCPAGQVK VLVDDGTCPAGQKVEVLGSK									
BRAO285v1_2940101	conserved protein of unknown function	6698	147 / 55,7	/	/	147	55,7	QMLEMHGPAIAQASQK ALSCSDGESEAREWR	Oxidation (M)								
BRAO285v1_2960006	conserved exported protein of unknown function	10525	201 / 41,8	/	/	201	41,8	QFVESGDDTIK VVLDSLTMSCK FGTYCAEHPVIVTAAEK	Oxidation (M)								
BRAO285v1_2960016	protein of unknown function	9215	56 / 13,3	/	/	56	13,3	MFGYNDVATGR	Oxidation (M)								
BRAO285v1_2980014	conserved protein of unknown function	15535	75 / 9,9	/	/	75	9,9	LGGVYTGVTEDQLR									
BRAO285v1_310002	putative Type III polyketide synthase	36220	/	55 / 2,6	55 / 2,6		55	2,6	MRPATRASR					55	2,6	MRPATRASR	
BRAO285v1_340006	conserved protein of unknown function	18871	70 / 8,4	/	/	70	8,4	LQMHLQETNQQQLQR	Oxidation (M)								
BRAO285v1_340020	conserved exported protein of unknown function	22686	/	/	104 / 8,3									104	8,3	AADDAQDKDVPVVAR MKAADDAQDKDVPVVAR	Deamidated (NQ) Oxidation (M)
BRAO285v1_360070	conserved protein of unknown function	7016	/	88 / 18,2 74 / 18,2	71 / 18,2		88 74	18,2 18,2	ALALGIGLGHQR ALALGIGLGHQR					71	18,2	ALALGIGLGHQR	
BRAO285v1_530001	conserved protein of unknown function	16055	69 / 6,3	/	/	69	6,3	FFTAQGLDNR									
BRAO285v1_570007	putative phosphoribosylformylglycinamide synthetase, PurS component	8255	51 / 13,6	424 / 93,8	298 / 64,2	51	13,6	SLGVDGVASVR		424	93,8	SLGVDGVASVR SLGVDGVASVRQGK AIEGALKSLGVDGVASVR VFDFELAGADKAAQEAALK AAADKLLANTVIENRYVELLG		298	64,2	SLGVDGVASVR VTVTLTKGLDPQK AIEGALKSLGVDGVASVR VFDFELAGADKAAQEAALK TGILDPQKAIAGALKSLGVDGVASVR	

Gene name	Protein function	Protein mass (Da)	ORS285 free-living state	Bacteroid A. indica ORS285	Bacteroid A. afraspera ORS285	ORS285 free-living state				Bacteroid A. indica - ORS285				Bacteroid A. afraspera - ORS285			
						Mascot Protein score	Protein coverage (%)	Peptide sequences	Peptide modifications	Mascot Protein score	Protein coverage (%)	Peptide sequences	Peptide modifications	Mascot Protein score	Protein coverage (%)	Peptide sequences	Peptide modifications
BRAO285v1_620078	HupA DNA-binding protein HU	10989	/	53 / 9,9	/					53	9,9	IATTTLSKR					
BRAO285v1_620141	conserved protein of unknown function, putative SUF system FeS cluster assembly associated domain	13495	54 / 5,6	/	/	54	5,6	LSGEIIA	Acetyl (N-term)								
BRAO285v1_650005	conserved protein of unknown function, putative CsbD-like domain	7066	961 / 97,2	736 / 97,2 181 / 48,6	59 / 23,6 341 / 79,2	961	97,2	QGIGEGATGNDR DAVNTAAGAANKNL LQEGEGAVQEVK MDKIKGATNEAVGK GKQQALGDAKQATK QATKDAVNTAAGAANK STMDKIKGATNEAVGK GSTMKIKGATNEAVGK QATKDAVNTAAGAANKNL QGIGEGATGNDRLQEGEGAVQEVK QGIGEGATGNDRLQEGEGAVQEVK AKQIGEGATGNDRLQEGEGAVQEVK	Gln->pyro-Glu (N-term Q) Deamidated (NQ) Oxidation (M)	736 181	97,2 48,6	LQEGEGAVQEVK DAVNTAAGAANKNL LQEGEGAVQEVK QATKDAVNTAAGAANK GSTMKIKGATNEAVGK QATKDAVNTAAGAANKNL QGIGEGATGNDRLQEGEGAVQEVK GKQQALGDAKQATKDAVNTAAGAANK QGIGEGATGNDRLQEGEGAVQEVK AKQIGEGATGNDRLQEGEGAVQEVK QATKDAVNTAAGAANK GSTMKIKGATNEAVGK QATKDAVNTAAGAANKNL	Gln->pyro-Glu (N-term Q) Oxidation (M) Gln->pyro-Glu (N-term Q) Oxidation (M)	59 341	23,6 79,2	GSTMKIKGATNEAVGK QATKDAVNTAAGAANK GSTMKIKGATNEAVGK QATKDAVNTAAGAANKNL QGIGEGATGNDRLQEGEGAVQEVK	Oxidation (M) Gln->pyro-Glu (N-term Q) Oxidation (M)
BRAO285v1_680003	Putative 4-oxalocrotonate tautomerase	7471	75 / 16,2	282 / 66,2 68 / 16,2	56 / 16,2 213 / 58,8	75	16,2	PEITINMAAGR	Oxidation (M)	282 68	66,2 16,2	PEITINMAAGR PEITINMAAGRTTEEK ALVTHLGVDPPEAVVIQINEAPLTHK DISKALVTHLGVDPPEAVVIQINEAPLTHK PEITINMAAGR	Oxidation (M) Oxidation (M)	56 213	16,2 58,8	PEITINMAAGR PEITINMAAGR ALVTHLGVDPPEAVVIQINEAPLTHK DISKALVTHLGVDPPEAVVIQINEAPLTHK	Oxidation (M) Oxidation (M)
BRAO285v1_70003	conserved protein of unknown function	7015	/	67 / 19,7	/					67	19,7	QGTHTGSPGVCR	Gln->pyro-Glu (N-term Q)				
BRAO285v1_710015	conserved protein of unknown function	8193	/	114 / 41,9	/					114	41,9	TTLLEQALQVQR LSADDQNAAGALLDYYK					
BRAO285v1_720027	conserved protein of unknown function	7400	50 / 15,9	/	/	50	15,9	TAMATPDNVLR	Oxidation (M)								
BRAO285v1_80019	rpmA 50S ribosomal subunit protein L27	9587	/	55 / 10	/					55	10	VIPGNIAR					
BRAO285v1_860016	conserved protein of unknown function	6431	72 / 21,7	121 / 53,3	119 / 46,7	72	21,7	AEPDDYIDAVGR		121	53,3	AEPDDYIDAVGR LVDEFSLADEAPASVFTA		119	46,7	AEPDDYIDAVGR ALSPLIEDSWRPAVR	
BRAO285v1_920004	thiS thiamine biosynthesis protein ThiS	7043	50 / 16,7	/	/	50	16,7	AGDEIITPR									
BRAO285v1_930001	putative ABC transporter, substrate-binding protein	44067	51 / 2,4	57 / 2,4	51 / 2,4	51	2,4	ASSPATASRR		57	2,4	ASSPATASRR		51	2,4	ASSPATASRR	Acetyl (N-term)
BRAO285v1_950060	pufB Light harvesting 1 beta subunit	8195	51 / 17,1	/	/	51	17,1	GGSLGLESEAR									
BRAO285v1_950109	conserved protein of unknown function	7161	/	59 / 20 54 / 16,9	/					59 54	20 16,9	GIVTYQLTSPLR PLTEDAIQVKG					
BRAO285v1_950136	bfr bacterioferritin (iron storage homoprotein)	18953	/	52 / 11,1	/					52	11,1	IGLELYQKHVGGLESEH					
BRAO285v1_990037	Putative ABC-type RTX toxin transporter, ATPase and permease components	80711	52 / 1,2	52 / 1,2 58 / 1,2	52 / 1,2 52 / 1,2	52	1,2	ASSPAIRRA	Acetyl (N-term)	52 58	1,2 1,2	ASSPAIRRA ASSPAIRRA	Acetyl (N-term) Acetyl (N-term)	52 52	1,2 1,2	ASSPAIRRA ASSPAIRRA	Acetyl (N-term) Acetyl (N-term)
back to table 1																	

Supplemental Table S4 : Accessions and origin of *Aeschynomene* species used in this study.

Species	Accession	Origin	Inoculated strain
<i>A. indica</i>	LSTM n°19	Senegal	ORS285
<i>A. evenia ssp.serrulata</i>	PI 572567	USA	ORS285
<i>A. evenia ssp. evenia</i>	PI 225551	Zambia	ORS285
<i>A. virginica</i>	LSTM n°63	USA	ORS285
<i>A. scabra</i>	PI 296044	Mexique	ORS285
<i>A. sensitiva</i>	LSTM n°28	Senegal	ORS285
<i>A. afraspera</i>	LSTM n° 1	Senegal	ORS285
<i>A. aspera</i>	IRRI 013020	Sri Lanka	ORS285
<i>A. nilotica</i>	IRRI 014040	Senegal	ORS285
<i>A. americana</i>	PI 544122	USA	ORS301
<i>A. pfundii</i>	LSTM n°54	Zimbabwe	ORS302
<i>A. schimperii</i>	IRRI 012156	Senegal	ORS305

Supplemental Table S5: List of primer pairs used in this study and their utilization. The same primer pair designed for *EFl α* was used for the 3 *Aeschynomene* species.

Gene	Forward primer	Reverse primer	Use
<i>EFlα</i>	AATGGTGATGCTGGTATGGTTAAG	TCTTCTTCTGTGCTGCCTTGG	RTqPCR
<i>AeDNF1</i>	TCTGCTCACCTTCGCCATAACC	ACCTCGTCATTGCCGTTGGG	RTqPCR
<i>AeDNF1</i>	CACCAGGTCAGATAGCGGTGTAAG	CAGAATGGTTATGGCGAAGG	RNAi
<i>AaNCR</i>	AAATTATTAGTTAAGATTCTCAC	ATCAAATGGGCATATACAAATG	ISH probe
<i>AiNCR</i>	TTGTCCACAGGAAAACCTTCTATTG	GCTACCGCCACACTCTCC	ISH probe
<i>Aa/AiLegHb</i>	TGGTGAACCTCATGGAATGTATTG	GAGATGGACAGAGCCTAAGTG	ISH probe
<i>AaLegHb</i>	CTCTTGTGGTGAACCTCATGGAATG	AATGGTGGCATGTGTCTTGAG	RTqPCR
<i>AaNCR1</i>	CAATGGCAAGTCATCATCTTCAAG	ATTTTCATCTAACGGGCACCTTTGG	RTqPCR
<i>AaNCR2</i>	ACAAGTCATCATCAAGTGGCAGAG	CAGGAACAGTCTTCACCATCAGG	RTqPCR
<i>AaNCR4</i>	ACTTATCTTGCTGGTGACAGTTATG	ATTTTCATCTAACGGGCACCTTTGG	RTqPCR
<i>AaNCR6</i>	GTATGGGCGGACACGAACTG	GGACTTGGGATATTGGGAGCAC	RTqPCR
<i>AiLegHb</i>	TAGGAGCCACCCATTTCAAATATG	CCCTTCATCTCTGCTGACCAC	RTqPCR
<i>AiNCR1</i>	CAGACAACATTTGCTACAAAGATG	AACCCTCGTCAAACCACATATC	RTqPCR
<i>AiNCR2</i>	TTGTCCACAGGAAAACCTTCTATTG	GCTACCGCCACACTCTCC	RTqPCR
<i>AiNCR3</i>	CATCATCAAGTGACTGGCATAACG	AGCCTAACGCAAACCTCCAAGC	RTqPCR
<i>AiNCR5</i>	GCCTCGGGAAAGTGTAACAA	TCTCTTGACATTGCCATCA	RTqPCR

Supplemental material and methods:

Proteins in the range of ~4-9 kDa (indicated with dashed line in Supplemental Figure S2) from free-living bacteria and bacteroid extracts were cut out from Tricine SDS-PAGE gels and destained with 500 μ L of 25 mM NH_4HCO_3 then twice with 50 % acetonitrile in 25 mM NH_4HCO_3 before being dried in a vacuum centrifuge system. Fifteen microliters of a trypsin solution (Sequencing Grade Modified Trypsin, Promega, Madison, USA) at a concentration of 0.0125 $\mu\text{g}/\mu\text{L}$ in 25 mM NH_4HCO_3 , was added to every gel slice which were kept for 15 min on ice. Ninety microliters of 25 mM NH_4HCO_3 were added, and the samples were kept another 15 min at room temperature. Digestion was performed overnight at 37°C then peptides were extracted twice by addition of 100 μL of 80 % acetonitrile in water containing 2 % formic acid. Peptides were then dried in a vacuum centrifuge before being suspended in 10 μL formic acid 2 % for LC-MS/MS analysis.

The protein digests were analyzed using a Q-TOF mass spectrometer (Maxis Impact; Bruker Daltonik GmbH, Bremen, Germany), interfaced with a nano-HPLC U3000 system (Thermo Scientific, Waltham, USA). Samples were concentrated with a pre-column (Thermo Scientific, C18 PepMap100, 300 $\mu\text{m} \times 5 \text{ mm}$, 5 μm , 100 Å) at a flow rate of 20 $\mu\text{L}/\text{min}$ using 0.1% formic acid. After pre-concentration, peptides were separated with a reversed-phase capillary column (Thermo Scientific, C18 PepMap100, 75 $\mu\text{m} \times 250 \text{ mm}$, 3 μm , 100 Å) at a flow rate of 0.3 $\mu\text{L}/\text{min}$ using a two steps gradient (8 % to 28 % acetonitrile in 40 min then 28 % to 42 % in 10 min), and eluted directly into the mass spectrometer. Proteins were identified by MS/MS by information-dependent acquisition of fragmentation spectra of multiple charged peptides. MS/MS raw data were analysed using Data Analysis software (Bruker Daltonik GmbH, Bremen, Germany) to generate the peak lists.

Protein identifications were obtained by searching locally with Mascot (v.2.4.0, Matrix Science) against a home-made nucleotide database (215376 sequences). This database was composed of *A. afraspera* and *A. indica* sequences from ESTtik databases (www.esttik.cirad.fr). All the NCRs were extracted from these libraries and their signal peptide sequences were manually removed. The following parameters were used: semi-trypsin as enzyme, two missed cleavage sites were allowed, carbamidomethylation of Cysteine as fixed modification and Acetyl (N-term), deamidated (NQ), Gln->pyro-Glu (N-term Q), Glu->pyro-Glu (N-term E), Oxidation (M), Phosphorylation (ST) and Phosphorylation (Y) were allowed as variable modifications. Mass tolerance was set to 30 ppm on full scans and 0.05 Da for fragment ions. Proteins were validated when at least one peptide with a p value <0.05 was identified. In addition, for all protein identified with only one peptide, MS/MS spectra were carefully checked manually for validation.

Supplemental material and methods:

Proteins in the range of ~4-9 kDa (indicated with dashed line in Supplemental Figure S2) from free-living bacteria and bacteroid extracts were cut out from Tricine SDS-PAGE gels and destained with 500 μ L of 25 mM NH_4HCO_3 then twice with 50 % acetonitrile in 25 mM NH_4HCO_3 before being dried in a vacuum centrifuge system. Fifteen microliters of a trypsin solution (Sequencing Grade Modified Trypsin, Promega, Madison, USA) at a concentration of 0.0125 $\mu\text{g}/\mu\text{L}$ in 25 mM NH_4HCO_3 , was added to every gel slice which were kept for 15 min on ice. Ninety microliters of 25 mM NH_4HCO_3 were added, and the samples were kept another 15 min at room temperature. Digestion was performed overnight at 37°C then peptides were extracted twice by addition of 100 μL of 80 % acetonitrile in water containing 2 % formic acid. Peptides were then dried in a vacuum centrifuge before being suspended in 10 μL formic acid 2 % for LC-MS/MS analysis.

The protein digests were analyzed using a Q-TOF mass spectrometer (Maxis Impact; Bruker Daltonik GmbH, Bremen, Germany), interfaced with a nano-HPLC U3000 system (Thermo Scientific, Waltham, USA). Samples were concentrated with a pre-column (Thermo Scientific, C18 PepMap100, 300 $\mu\text{m} \times 5 \text{ mm}$, 5 μm , 100 \AA) at a flow rate of 20 $\mu\text{L}/\text{min}$ using 0.1% formic acid. After pre-concentration, peptides were separated with a reversed-phase capillary column (Thermo Scientific, C18 PepMap100, 75 $\mu\text{m} \times 250 \text{ mm}$, 3 μm , 100 \AA) at a flow rate of 0.3 $\mu\text{L}/\text{min}$ using a two steps gradient (8 % to 28 % acetonitrile in 40 min then 28 % to 42 % in 10 min), and eluted directly into the mass spectrometer. Proteins were identified by MS/MS by information-dependent acquisition of fragmentation spectra of multiple charged peptides. MS/MS raw data were analysed using Data Analysis software (Bruker Daltonik GmbH, Bremen, Germany) to generate the peak lists.

Protein identifications were obtained by searching locally with Mascot (v.2.4.0, Matrix Science) against a home-made nucleotide database (215376 sequences). This database was composed of *A. afraspera* and *A. indica* sequences from ESTtik databases (www.esttik.cirad.fr). All the NCRs were extracted from these libraries and their signal peptide sequences were manually removed. The following parameters were used: semi-trypsin as enzyme, two missed cleavage sites were allowed, carbamidomethylation of Cysteine as fixed modification and Acetyl (N-term), deamidated (NQ), Gln->pyro-Glu (N-term Q), Glu->pyro-Glu (N-term E), Oxidation (M), Phosphorylation (ST) and Phosphorylation (Y) were allowed as variable modifications. Mass tolerance was set to 30 ppm on full scans and 0.05 Da for fragment ions. Proteins were validated when at least one peptide with a p value <0.05 was identified. In addition, for all protein identified with only one peptide, MS/MS spectra were carefully checked manually for validation.