nature biotechnology

Genome sequence and identification of candidate vaccine antigens from the animal pathogen *Dichelobacter nodosus*

Garry S A Myers[1,8], Dane Parker[2–4,8], Keith Al-Hasani[2,3], Ruth M Kennan[2,3], Torsten Seemann[4], Qinghu Ren[1], Jonathan H Badger[1], Jeremy D Selengut[1], Robert T DeBoy[1], Hervé Tettelin[1], John D Boyce[2,3], Victoria P McCarl[2,5], Xiaoyan Han[2,3], William C Nelson[1], Ramana Madupu[1], Yasmin Mohamoud[1], Tara Holley[1], Nadia Fedorova[1], Hoda Khouri[1], Steven P Bottomley[2,5], Richard J Whittington[2,6], Ben Adler[2–4], J Glenn Songer[5], Julian I Rood[2–4] & Ian T Paulsen[1]

Dichelobacter nodosus **causes ovine footrot, a disease that leads to severe economic losses in the wool and meat industries. We sequenced its 1.4-Mb genome, the smallest known genome of an anaerobe. It differs markedly from small genomes of intracellular bacteria, retaining greater biosynthetic capabilities and lacking any evidence of extensive ongoing genome reduction. Comparative genomic microarray studies and bioinformatic analysis suggested that, despite its small size, almost 20% of the genome is derived from lateral gene transfer. Most of these regions seem to be associated with virulence. Metabolic reconstruction indicated unsuspected capabilities, including carbohydrate utilization, electron transfer and several aerobic pathways. Global transcriptional profiling and bioinformatic analysis enabled the prediction of virulence factors and cell surface proteins. Screening of these proteins against ovine antisera identified eight immunogenic proteins that are candidate antigens for a cross-protective vaccine.**

D. nodosus is a Gram-negative aerotolerant anaerobe that is the principal causative agent of footrot in ruminants[1]. In sheep, footrot is a highly infectious and debilitating disease that results in lameness, weight loss, failure to thrive and poor wool growth[2,3]. Ovine footrot begins as an interdigital dermatitis and progresses to destruction of the epidermal matrix and necrotic separation of the hoof from the underlying soft tissue[3]. Infection with *D. nodosus* requires climatic conditions of warm weather and moist pastures. Treatment involves a combination of methods, including relocation to a dry paddock, foot-bathing, antibiotic therapy and vaccination, at considerable economic cost.

D. nodosus is the first member of the order Cardiobacteriales (γ-Proteobacteria)[4] to be sequenced. Other members of this poorly characterized order include the fastidious opportunistic pathogens *Cardiobacterium hominis* and *C. valvarum*, which are associated with infectious endocarditis[5], and *Suttonella* spp., which cause human eye infections, infectious endocarditis and mortality in wild birds[6].

We sequenced *D. nodosus* strain VCS1703A because it is both virulent and naturally transformable and has been used in footrot virulence studies[7,8]. Our analysis indicates that *D. nodosus* is unusual in having a compact genome with no discernible evidence for ongoing genome reduction. However, bioinformatic and comparative genome hybridization (CGH) data suggest the occurrence of extensive horizontal gene transfer. We identified vaccine candidates that were overexpressed, purified and screened against ovine antisera, permitting the discovery of candidate antigens for a cross-protective vaccine.

RESULTS

Genome features

The *D. nodosus* VCS1703A genome consists of a single circular chromosome of 1,389,350 base pairs (bp) (**Fig. 1**) with 1,299 predicted genes (**Table 1**). *D. nodosus* is located in the γ-Proteobacteria on the basis of 16S rRNA phylogeny[4]. Phylogenetic analysis of all *D. nodosus* genes (http://www.tigr.org/gdn) revealed that 57% clustered with γ-Proteobacteria, 13% clustered with β-Proteobacteria and 30% clustered with other taxonomic groupings (**Supplementary Fig. 1** online), confirming that *D. nodosus* is a deep-branching member of the γ-Proteobacteria (**Supplementary Fig. 2** online).

The small size of the *D. nodosus* genome makes it the smallest anaerobic bacterium sequenced to date and places it among the smallest non-intracellular bacterial pathogens yet characterized. The recently sequenced oceanic α-Proteobacterium *Pelagibacter ubique*

[1]The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, Maryland 20850, USA. [2]Australian Research Council Centre of Excellence in Structural and Functional Microbial Genomics, [3]Department of Microbiology, [4]Victorian Bioinformatics Consortium and [5]Department of Biochemistry and Molecular Biology, Monash University, Melbourne, Victoria 3800, Australia. [6]Faculty of Veterinary Science, University of Sydney, Sydney, New South Wales 2570, Australia. [7]Department of Veterinary Science, University of Arizona, Tucson, Arizona 85721, USA. [8]These authors contributed equally to this work. Correspondence should be addressed to I.T.P. (ipaulsen@tigr.org).

Received 20 November 2006; accepted 4 April 2007; published online 29 April 2007; doi:10.1038/nbt1302

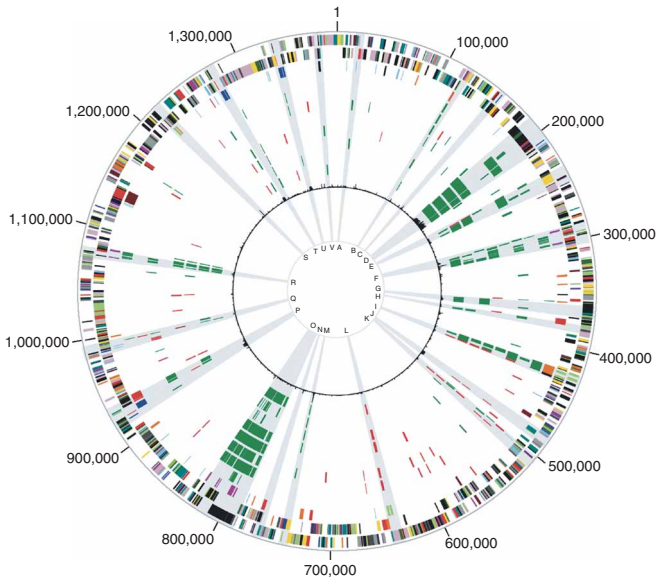

Figure 1 Circular representation of the *D. nodosus* VCS1703A genome. Data are described from outermost circle to innermost. Circles 1 and 2: tick marks represent predicted coding sequences on the plus strand (circle 1) and minus strand (circle 2) colored by cellular role. Role categories and colors are as follows: amino acid biosynthesis, violet; biosynthesis of cofactors, prosthetic groups and carriers, light blue; cell envelope, light green; cellular processes, red; central intermediary metabolism, brown; DNA metabolism, gold; energy metabolism, light gray; fatty acid and phospholipid metabolism, magenta; protein synthesis and fate, pink; biosynthesis of purines, pyrimidines, nucleosides and nucleotides, orange; regulatory functions and signal transduction, olive; transcription, dark green; transport and binding proteins, blue-green; other categories, salmon; unknown function, gray; conserved hypothetical proteins, blue; hypothetical proteins, black. Circle 3: coding sequences of interest are represented as follows: bacteriophage genes, magenta; IS elements, black; rRNA, blue; tRNA, light blue; outer membrane protein (OMP) genes, orange; protease genes, light green; fimbrial structural genes, red; fimbrial regulatory genes, brown. Circles 4–11: genomic microarray data for the following *D. nodosus* strains: circle 4, VCS1690A; circle 5, HA320; circle 6, AC424; circle 7, A198; circle 8, CS101; circle 9, AC390; circle 10, 806(2a); circle 11, 375(4b). Genes missing or divergent relative to VCS1703A, green; gene intensity value amplified relative to VCS1703A, red. Circle 12: χ^2 analysis of trinucleotide composition. Light blue overlays, labeled A–V (clockwise), represent regions of interest identified through CGH and bioinformatic analysis (see text).

currently has the smallest known genome of a free-living organism known to replicate independently in nature[9]. *P. ubique* (1,308,759 bp) does have a slightly smaller genome than *D. nodosus*; however, *D. nodosus* has fewer genes (1,299, as compared to 1,354 in *P. ubique*) owing to its larger intergenic spacing.

Most organisms with relatively small genomes are obligate intracellular pathogens or symbionts with evidence of extensive genome reduction[10]. The genomic organization of *D. nodosus*, like that of *P. ubique*, differs from those of obligate intracellular organisms. For example, *Rickettsia prowazekii* and *Mycobacterium leprae* have numerous pseudogenes, indicative of ongoing genome reduction. In contrast, only six pseudogenes have been identified in *D. nodosus*. There is little evidence of gene-family expansion in *D. nodosus*; the largest family of recent paralogous gene duplications comprises four genes encoding the major outer membrane proteins (see below). Intracellular organisms with small genome sizes typically have highly reduced metabolic capabilities and a limited transport repertoire. Both *D. nodosus* and *P. ubique* have greater biosynthetic capabilities than obligate intracellular organisms with similar genome sizes (see below). The short intergenic spacing (averaging 3 bp) observed in *P. ubique* may result from "genome streamlining"[9]. By contrast, *D. nodosus* shows an average intergenic spacing of 84 bp, close to the median observed in prokaryotic genomes, indicating that the small genome sizes of *D. nodosus* and *P. ubique* probably result from differing evolutionary pressures exerted by their respective environmental niches and effective population sizes.

Lateral transfer, virulence and diversity

D. nodosus differs from other organisms with small genomes in showing a high degree of apparent lateral gene transfer. This was assessed by analysis of regions of atypical trinucleotide composition, combined with CGH microarray experiments on eight *D. nodosus* isolates (**Supplementary Table 1** online). Although atypical nucleotide composition is a simplistic measure of lateral gene transfer, it can point to potential acquisitions, particularly in combination with other evidence. Nineteen percent of *D. nodosus* VCS1703A genes were located in regions of atypical trinucleotide composition and were variable in strains as assessed by CGH, suggesting recent lateral acquisition. Analysis of the phylogenetic lineages of genes in these regions indicated that 65% clustered outside of the γ-Proteobacteria, a percentage higher than that observed for the overall genome. Twenty-one distinct zones of unusual trinucleotide composition were identified (regions A–V; **Fig. 1**; **Supplementary Fig. 3** online); most also corresponded to strain-specific variable regions based on CGH data.

The largest region of atypical trinucleotide composition is a 38.4-kb integrated Mu-like bacteriophage (B. Cheetham, personal communication; zone O, **Fig. 1**). This region encodes homologs of phage terminase and head and tail morphogenesis proteins. This region is divergent in at least five of the eight isolates examined (confirmed by Southern blotting; data not shown). *attP*-like sites and phage-like integrases have been reported to be associated with *D. nodosus* genomic islands[11], but a complete *D. nodosus* bacteriophage-like sequence has not been reported previously.

Several facets of *D. nodosus* virulence have been previously identified, including type IV fimbriae[8], extracellular proteases[12] and outer membrane proteins[13]. Additionally, two types of genomic islands have been associated with virulent *D. nodosus* isolates: the *vap* and *vrl* regions[12,14,15]. Two *vap* loci that include phage-like integrases and plasmid gene homologs are present in *D. nodosus* VCS1703A,

Table 1 General features of the *D. nodosus* genome

	D. nodosus chromosome
Size (base pairs)	1,389,350
G+C content (%)	44.4
Protein coding genes:	
Similar to known proteins	1,032
Conserved hypothetical	131
Hypothetical	136
Total	1,299
Average gene size (base pairs)	973
Coding (%)	91.5
rRNA operons	3
tRNA genes	45
Structural RNA	1

corresponding to the unusual trinucleotide composition zones F and R (**Fig. 1**). These *vap* loci are conserved in two *D. nodosus* isolates (HA320, with intermediate virulence; and the virulent strain A198), with the remainder of the isolates showing variability. Similarly, VCS1703A contains the *vrl* island (region D; **Fig. 1**), also associated with unusual trinucleotide composition. The *vrl* island from strain VCS1703A has only 3 bp that are different from the 27-kb *vrl* region of strain A198 (ref. 12). The *vrl* region was also present in strain HA320 and AC424, a benign isolate. Despite the association of *vap* and *vrl* loci with *D. nodosus* virulence, no *vap* or *vrl* genes have a known virulence function. Loci homologous to the *D. nodosus vrl* region are found in various environmental eubacteria from distinct clades, such as *Desulfovibrio*, *Nitrosococcus* and *Chlorobium*, suggesting that *vrl* has been transferred laterally between diverse organisms.

Type IV fimbriae are present in many important pathogens; in *D. nodosus* they are required for virulence, natural transformation, adhesion, twitching motility and protease secretion[8,16]. The fimbriae are used to classify *D. nodosus* strains into ten serogroups; vaccination using the fimbriae is effective in preventing homologous but not heterologous challenge[3,17]. The type IV fimbrial biogenesis genes are scattered throughout the genome in eight locations. The *fimAB* genes are in a region of atypical trinucleotide composition (zone B, **Fig. 1**), and microarray analysis indicates that *fimA* is divergent in all but two strains. *fimA* encodes the type IV fimbrial subunit; variation in *fimA* is the primary basis for *D. nodosus* serogroup typing[17]. The two strains with *fimA* conserved belong to serogroups G and I; the sequenced strain belongs to serogroup G. *fimB* is divergent or missing in four strains, three of which are type II strains that characteristically lack *fimB*[18].

The genome contains 21 putative fimbrial biogenesis genes and 10 fimbrial regulatory genes, whereas there are 43 fimbrial genes in *P. aeruginosa* (**Supplementary Fig. 4** online). Many of the missing genes are involved in regulation. Genes potentially involved in twitching motility (*pilT*, *pilU*) and its regulation (*chpA*, *pilJ*, *pilI*, *pilG*, *pilH*, *ppk*, *fimX*) were identified in addition to putative tip adhesin (*pilC*) and secretin (*pilQ*) genes. *pilQ* was located in a region of atypical nucleotide composition (zone H, **Fig. 1**) that is variable based on microarray and PCR analysis.

No genes encoding a separate type II secretion system were identified in *D. nodosus*. However, several putative type IV fimbrial biogenesis genes showed similarity to type II secretion genes. Several of the type IV fimbrial genes were putative alternative pilin genes (**Supplementary Fig. 3**), although there were no type II secretion–specific pseudopilin or *gspC* genes. The fimbrial subunit is involved in protease secretion in *D. nodosus*[8,16], and PilQ seems to be the only outer membrane secretin encoded by *D. nodosus*. Given that some type II secretion proteins share similarity with fimbrial proteins, the type IV fimbriae may be a secretion portal, expelling proteins through the motive force of fimbrial extension and retraction[19]. We propose that in *D. nodosus* the type IV fimbrial secretin PilQ may act with the fimbrial biogenesis machinery as the only type II protein secretion system.

D. nodosus secretes three closely related extracellular proteases[20], most likely involved in invasion and penetration by digestion of the hoof epidermal matrix. Virulent isolates produce two acidic (AprV2, AprV5) proteases and one basic (BprV) protease that belong to the subtilisin protease family and are encoded within regions of unusual trinucleotide composition (AprV5 and BprV, zone L; AprV2, zone S; **Fig. 1**).

The major outer membrane protein Omp1 of *D. nodosus* is phase variable, switching between four closely related and clustered genes (**Supplementary Fig. 5** online)[13]. This locus is located in a region of unusual trinucleotide composition (zone I; **Fig. 1**) that is highly divergent in all strains as determined by CGH. This four-gene cluster is flanked on one side by two insertion sequence (IS) elements, belonging to the IS*200* and IS*605* families. These transposase genes were the only genes clearly present in multiple copies in some strains as determined by CGH.

We identified virulence factors associated with regions of atypical nucleotide composition. Zone G encodes a putative secreted repeats-in-toxin (RTX)-like protein (DNO_0334) and two ABC-family efflux proteins (DNO_0335-0336). RTX proteins are large pore-forming toxins that damage host cells and impair host defense mechanisms, and that are important virulence factors in various Gram-negative bacteria[21]. A *D. nodosus* RTX-like toxin could play a role in the necrotic sequelae of footrot. Another region of atypical nucleotide composition (zone M; **Fig. 1**) encodes a putative large, repetitive secreted protein (DNO_0690; 32 nine-amino-acid repeat units) that may mediate adhesion to the extracellular matrix. This protein could be associated with virulence, as it is conserved in all virulent strains and absent in all benign strains (**Fig. 1**).

Several other genes are homologous to virulence-associated genes in other organisms, or to genes encoding protective antigens. These included two genes (DNO_0466 and DNO_0650) homologous to *Shigella flexneri ispA* and *vacJ* genes that are involved in inter- and intracellular spreading[22]. A homolog (DNO_1067) of cell wall–associated hydrolases was also identified. This protein showed similarity to a putative vaccine candidate (GNA2001) from *N. meningitidis*[23], and to *Listeria monocytogenes* cell surface protein P60, which promotes phagocyte invasion[24]. Another gene (DNO_0681) showed similarity to the highly immunogenic D15 and Oma87 proteins from *Haemophilus influenzae* and *Pasteurella multocida*, respectively, which mediate protection in animal models[25]. We also identified a serine protease (DNO_0902) similar to *P. aeruginosa* MucD, which is involved in alginate biosynthesis and extracellular toxin production[26].

A type I restriction/modification system (DNO_0222-0223) is encoded within another region of atypical nucleotide composition (zone E; **Fig. 1**), consistent with the low natural transformation efficiency of strain VCS1703A[7]. This locus is missing or divergent in several *D. nodosus* strains and represents a target for mutagenesis to improve VCS1703A transformability.

The regulatory capacity of an organism is correlated with genome size across a swathe of different bacteria[27]. In *D. nodosus*, we identified a modest suite of regulatory genes encoding 21 transcription factors, three sigma factors (RpoD, RpoH, RpoN), seven sensor kinases and nine response regulators. Five two-component regulatory systems were identified, which were predicted to be involved in functions including chemosensing and phosphate utilization. One system, PilSR, is required for the regulation of the type IV fimbriae[16]. Only 3% of the *D. nodosus* genome is involved in regulation, as compared to 8.4% in *P. aeruginosa* and 8% in *Escherichia coli*[28,29], reflecting the small genome and specialized niche of *D. nodosus*.

Metabolism and transport

Metabolic reconstruction of *D. nodosus* (**Fig. 2**) predicted various capabilities not suspected from the early literature. No minimal medium exists for *D. nodosus*; however, it has been thought to be incapable of growth on carbohydrates, relying on amino acid fermentation for carbon and energy[4]. Nevertheless, we identified components of carbohydrate metabolism in the genome: a partial glycolytic pathway (fructose 1,6-biphosphate to pyruvate) and a complete non-oxidative branch of the pentose phosphate pathway. Both fructose and glycerol can probably be fed into glycolysis via a fructose PTS transporter and phosphofructokinase, and a glycerol MIP channel and glycerol catabolic pathway. However, growth experiments using

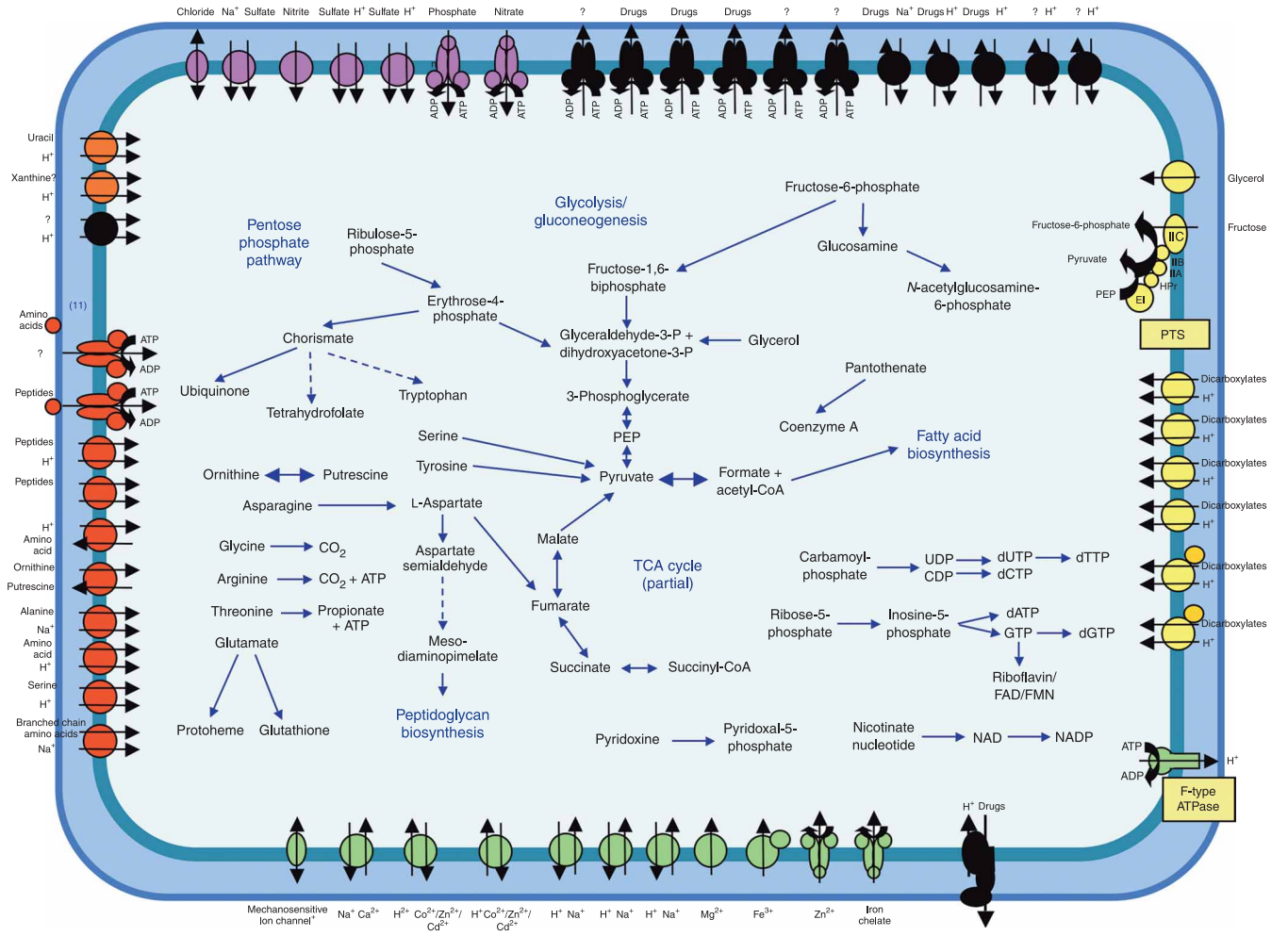

Figure 2 Overview of metabolism and transport in *D. nodosus*. Dotted arrows for metabolic pathways indicate that the complete pathway was not identified. Transporters are grouped by substrate specificity and indicated by color as follows: inorganic cations (green), inorganic anions (magenta), carbohydrates and carboxylates (yellow), amino acids and peptides (red), nucleosides and nucleobases (orange) and drug efflux and other (black). Arrows indicate direction of transport for substrates (and coupling ions, where appropriate).

fructose and glycerol as the carbon sources did not enhance the naturally poor growth of *D. nodosus* (data not shown). A partial trichloroacetic acid (TCA) cycle (succinyl-CoA to malate) was identified, and at least six carboxylate transporters are present. Similarly, *D. nodosus* was thought to be incapable of electron transport. However, metabolic reconstruction identified a cytochrome (quinine) oxidase, NADH (ubiquinone) oxidoreductase, fumarate reductase and a complete ubiquinone biosynthesis pathway.

D. nodosus has been described as an aerotolerant anaerobe[4]. Putative oxidative stress genes encoding a manganese superoxide dismutase, alkyl hydroperoxide reductase and methionine sulfoxide reductase were identified, although no catalase or peroxidase genes were present. Notably, genes encoding aerobic metabolic capabilities were identified, including an aerobic ribonucleotide reductase, aerobic ubiquinone biosynthesis pathway, cytochrome (quinine) oxidase and miscellaneous oxidases. This suggested that *D. nodosus* might grow microaerophilically. We tested this by growth experiments using *Campylobacter* GasPaks; *D. nodosus* was capable only of weak short-term, but not sustained, growth under microaerophilic conditions. When anaerobically grown cells were subcultured onto blood agar plates and then incubated under microaerophilic conditions, minimal growth was observed on the primary streak. However, if these cells were subcultured a second time under microaerophilic conditions, no growth was detected even though the same cells were capable of growth under anaerobic conditions. We investigated the ability of *D. nodosus* to tolerate aerobic and microaerophilic conditions. *D. nodosus* cells on agar plates remained viable after 10 d of exposure to aerobic conditions. This suggests a scenario whereby *D. nodosus* subjected to periodic exposure to oxygen in the ruminant hoof remains viable.

Intracellular pathogens and symbionts with small genome sizes tend to have lost many biosynthetic pathways, although the 'streamlined' free-living *P. ubique* has more robust biosynthetic capabilities than expected from its genome size[9,10]. **Supplementary Table 2** online provides a comparison of biosynthetic pathways in organisms with small genome sizes. *D. nodosus* encodes pathways absent in many intracellular organisms, including the NAD, glutathione, pantothenate, riboflavin and folate biosynthesis pathways. It lacks various metabolic pathways present in *P. ubique*, particularly those related to amino acid biosynthesis.

Because of this paucity of amino acid biosynthetic pathways, amino acid transport provides an excellent example of genome efficiency in *D. nodosus*. Amino acid ABC transporters typically consist of several

Table 2 Differentially expressed genes of *D. nodosus* when grown on hoof agar, validated with QRT-PCR

	Microarray		QRT-PCR		
Gene	Hoof/control ratio	*P* value	Hoof/control ratio	s.d.	*P* value
feoA	0.54	1.81×10^{-5}	0.40	0.15	1.46×10^{-2}
pilU	2.07	8.43×10^{-7}	5.84	2.41	4.73×10^{-4}
znuA	5.06	3.81×10^{-6}	16.4	4.76	2.16×10^{-5}
ihfA	0.32	4.88×10^{-4}	0.21	0.12	7.88×10^{-5}
hslU	2.26	1.19×10^{-7}	1.72	0.72	6.92×10^{-3}

genes encoding ATP-binding proteins, membrane proteins and a periplasmic binding protein. *D. nodosus* encodes a single set of ATP-binding and membrane proteins possibly involved in amino acid transport. However, there are 11 amino acid binding protein genes scattered around the genome that presumably provide *D. nodosus* with transport specificities for different amino acids while minimizing gene usage. Although the three extracellular proteases noted previously may be virulence factors, a more critical function may be digesting host proteins into their constituent amino acids, to be used directly for protein synthesis or as carbon and energy sources.

In vitro expression of *D. nodosus* genes

Little is known regarding *D. nodosus* gene expression in the ovine hoof. *D. nodosus* populations are probably not high enough to permit microarray analysis using *in vivo* material. To model *in vivo* conditions and examine changes in gene expression, we extracted RNA from cells grown on Eugon yeast extract agar and on the same medium containing 2% ground ovine hoof powder. We identified 86 genes that were differentially expressed (**Supplementary Table 3** online), covering a spectrum of functions congruent with a growth-change experiment. Downregulation of the global regulator integration host factor (IHF) was observed; conversely, increased expression was detected for the twitching motility gene *pilU* and a zinc uptake gene *znuA*. These and several other genes were validated by quantitative real-time RT-PCR (QRT-PCR; **Table 2**; **Supplementary Table 4** online). These results point to genes that are potentially expressed upon contact of *D. nodosus* with the ovine hoof.

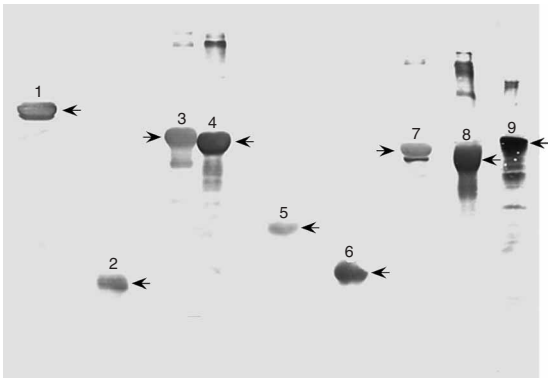

Figure 3 Immunoblot demonstrating recognition of *D. nodosus* proteins by pooled sera from experimentally infected sheep. Protein bands: 1, DNO_0012 (37 kDa + 60 kDa NusA tag); 2, DNO_0033 (29 kDa); 3, (67 kDa); 4, DNO_0605 (64 kDa); 5, DNO_0644 (34 kDa); 6, DNO_0725 (29 kDa); 7, DNO_1167 (66 kDa); 8, DNO_1241 (57 kDa); 9, *D. nodosus* whole cell lysate.

Table 3 Proteins showing specific reactivity to *D. nodosus* immune sera

Gene name	Function
DNO_0012	Peptidyl-prolyl *cis-trans* isomerase, FKBP-type
DNO_0033	Bacterial extracellular solute-binding–family protein
DNO_0603	Acidic extracellular subtilisin-like protease precursor AprV5
DNO_0605	Basic extracellular subtilisin-like protease precursor BprV
DNO_0644	Periplasmic iron-binding protein YfeA
DNO_0725	Potential adhesin complex protein
DNO_1167	Acidic extracellular subtilisin-like protease AprV2
DNO_1241	Hypothetical lipoprotein

FKBP, FK506 binding protein.

Identification of vaccine candidates

Vaccination against *D. nodosus* using whole cells or purified fimbriae has been available for some time[30], but its value is limited by the need for multivalent vaccines and by antigenic competition problems associated with such vaccines[2]. *D. nodosus* can undergo serogroup conversion, and reservoirs of serogroups are found in infected hooves, highlighting the need for heterologous protection[17,31,32]. Poor cross-protection by vaccines composed of whole or depiliated cells indicate that other cross-protective factors exist[33,34]. The development of an effective vaccine against all serogroups requires the identification of these antigens and would be of major benefit to the wool and meat industries.

We used high-throughput reverse vaccinology to identify cross-protective candidates. A bioinformatic screen identified 99 predicted surface-exposed or secreted protein candidates; 87 were cloned into hexahistidine and/or NusA-tagged expression vectors. Recombinant proteins were screened against pooled immune and preimmune sheep sera using western blotting, leading to the identification of eight proteins that reacted with the immune, but not preimmune, sera (**Table 3**; **Fig. 3**; **Supplementary Fig. 6** online).

Three of these eight (**Table 3**) were the extracellular proteases, which have a predicted role in virulence via digestion of the hoof, aiding bacterial infiltration. Others include the Fur-regulated iron binding protein YfeA, which has been shown to be secreted[35], and a peptidyl-prolyl *cis-trans* isomerase (DNO_0012) homologous to the *Legionella pneumophila* macrophage infectivity potentiator (MIP), an essential cell-surface virulence factor[36]. Similarly, a *Neisseria gonorrhoeae* MIP homolog is required for intramacrophage survival[37]. Immunological reactivity to unique antigens alone does not prove a role in protective immunity against footrot. Nevertheless, identification of these antigens may aid the development of an efficacious cross-protective vaccine.

DISCUSSION

Sequencing and analysis of the *D. nodosus* genome has revealed a surprising degree of strain diversity and lateral gene acquisition for an organism with a small genome. A large array of biosynthetic capabilities is present for such a small genome, yet the genome lacks amino acid processing pathways. Determination of the complete genome sequence of *D. nodosus* has permitted the application of high-throughput approaches, such as global transcriptional profiling and reverse vaccinology, to understanding ovine footrot.

METHODS

Growth conditions. *D. nodosus* strains were grown on Eugon (Difco) yeast extract (EYE) agar containing 5% (vol/vol) defibrinated horse blood (Equicell)

or in Eugon (Difco) broth with yeast extract in an anaerobic chamber as described[35]. Anaerobic jar experiments were performed using AnaeroGen and CampyGen atmosphere-generation sachets (Oxoid) in Oxoid anaerobic jars.

Genome sequencing and annotation. The complete genome sequence of the virulent serogroup G *D. nodosus* strain VCS1703A was determined using the whole-genome shotgun method as previously described[38]. Physical and sequencing gaps were closed using a combination of primer walking, generation and sequencing of transposon-tagged libraries of large-insert clones, and multiplex PCR[39]. Identification of putative protein-encoding genes and annotation of the genome were done as previously described[40]. An initial set of genes predicted to encode proteins was identified with GLIMMER[41]. Genes consisting of fewer than 30 codons and those containing overlaps were eliminated. Frameshifts and point mutations were corrected or designated 'authentic.' Functional assignment, identification of membrane-spanning domains, determination of paralogous gene families and identification of regions of unusual nucleotide composition were done as previously described[38–40]. Sequence alignments and phylogenetic trees were generated using methods previously described[38].

Trinucleotide composition. Distribution of all 64 trinucleotides (trimers) was determined, and the trimer distribution in 1,000-bp windows that overlapped by half their length (500 bp) across the genome was computed. For each window, we computed the χ^2 statistic for the difference between its trimer content and that of the whole chromosome. A large value of χ^2 indicates the trimer composition in this window is different from the rest of the chromosome (minimum of 2 s.d.; **Supplementary Table 5** online). Probability values for this analysis are based on assumptions that the DNA composition is relatively uniform throughout the genome, and that trimer composition is independent. Because these assumptions may be incorrect, we prefer to interpret high χ^2 values as indicators of regions on the chromosome that appear unusual and demand further scrutiny.

Comparative genomics. The *D. nodosus* VCS1703A genome was compared to other genomes at the nucleotide level by suffix tree analysis using MUMmer. *D. nodosus* CDSs were compared by BLAST against the complete set of non-redundant CDSs using an *E*-value cutoff of 1×10^{-5}.

The Automated Phylogenetic Inference System (APIS) was used for comparative phylogenetic analyses. APIS is an internal system of The Institute for Genomic Research (TIGR) for automatic creation and summarizing of phylogenetic trees for each protein encoded by a genome. The homologs used by APIS for each phylogenetic tree are obtained by comparing each query protein against a curated database of proteins from complete genomes using WU-BLAST (http://blast.wustl.edu). The full-length sequences of these homologs are then retrieved from the database and aligned using MUSCLE[42], and bootstrapped neighbor-joining trees are produced using QuickTree[43]. As QuickTree (unlike most programs) produces bootstrapped trees with meaningful branch lengths, the inferred tree is then midpoint rooted before analysis, allowing automatic determination of the taxonomic classification of the organisms with proteins in the same clade as the query protein. Note that bootstrap values are not used as a threshold—rather APIS uses bootstrapped trees in order to increase the accuracy of topology; the clades in such trees are the consensus of those in the individual bootstrap replicates. The large number of sequences used in generating each tree should also reduce the occurrence of artifacts due to long branch attraction, as the potential long branches would tend to be broken up by the presence of closely related sequences.

Comparative genomic microarray analysis. Genomic DNA was extracted from 2-d-old EYE agar cultures of *D. nodosus* using the DNeasy extraction kit (Qiagen). DNA (4 μg) was digested with 20 U of *Alu*I for 2 h at 37 °C and purified (PCR purification kit, Qiagen). Labeled genomic DNA (2 μg in 20 μl) was prepared by the addition of 25 μg of random hexamers with 20 μl of reaction buffer (42 mM 2-mercaptoethanol, 21 mM $MgCl_2$, 210 mM Tris-HCl, pH 7.0), boiled for 5 min and then placed on ice. Fluorescent dyes (60 nM; Cy5-dUTP or Cy3-dUTP, Amersham Biosciences) were coupled using 40 U of Klenow enzyme and nucleotides (1.2 mM dCTP, dGTP and dATP and 0.6 mM dTTP) in a final volume of 50 μl. Reactions were incubated for 2 h and stopped in the presence of 50 mM EDTA, pH 8.0. Reaction mixtures were purified with Microcon columns (Millipore) and concentrated in a Speedvac SVC (Savant). Samples were then hybridized, washed, scanned and analyzed as before using two slides per strain including a dye swap[16]. Genes that showed greater than a 1.5-fold change in signal and had a Wilcoxon signed-ranked *P* value <0.05 were included in the final dataset.

Transcriptional profiling. RNA was extracted as before[16] from 2-d-old EYE agar cultures of *D. nodosus* with and without 2% (wt/vol) ground hoof[44]. RNA (3 μg) was labeled using the 3DNA Array 900MPX labeling kit (Genisphere) and hybridizations were carried out as before[16], using the enhanced hybridization buffer at 52 °C for cDNA hybridization and the SDS hybridization buffer at 48 °C for 3DNA hybridization. Slides were scanned and analyzed as before[16] on four biological repeat samples with two dye swaps. QRT-PCR was done on an ABI PRISM 7700 sequence detector as described[16] using four biological replicates performed in triplicate (**Table 2**). Statistical analysis was performed using a two-tailed Student's *t*-test. Microarray data have been deposited at the National Center for Biotechnology Information's Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo; see accession codes below).

Vaccine candidate selection and analysis. Putative vaccine candidates were identified from the *D. nodosus* genome based on predicted cell localization and the presence of signal and lipoprotein peptides using the programs PSORTB[45], SignalP[46] and LipoP[47] and the presence of less than two predicted transmembrane domains. Genes were cloned into the expression vectors (pBAD-DEST49, pDEST-17 and a Gateway-adapted expression vector containing a NusA solubility tag, pLIC-Nus[48]) lacking signal peptides using the Gateway recombination system (Invitrogen).

The recombinant expression clones were assessed for expression and solubility using an inclusion-body assay on a TECAN liquid handling robot. Briefly, 1 ml induced cultures (Overnight Express, Merck) were lysed with PopCulture and Lysonase (Merck) for 20 min at 25 °C in a deep 24-well plate. Cell lysate (1 ml) was then added to a 96-well filter plate (AcroPrep) and the solution was drawn through the filter under vacuum. The inclusion bodies were retained while soluble proteins passed through the filter. The retained inclusion bodies were washed once with Triton X-100, to remove any remaining soluble proteins, and then twice with sodium phosphate–buffered saline (PBS), pH 7.4. The washed proteins were then denatured by the addition of 200 μl of 8 M urea to each corresponding well, incubated for 2 h at 25 °C and collected under vacuum. Both soluble and insoluble fractions were then run on a SDS-PAGE gel to assess the solubility of each protein. The recombinant expression clones were shown to express measurable levels of protein (0.2–3 mg/ml) in *E. coli* when analyzed by Coomassie staining. In general, protein expression levels were substantially higher when the recombinant protein targets were fused with a NusA solubility tag. All of the recombinant proteins were screened against pooled sera from five infected (strain VCS1001, serogroup A) sheep and five preimmune sheep by western blotting. Fractions containing comparable levels of protein were separated by SDS-PAGE, transferred to nitrocellulose membranes and incubated with sera (1:200) before the addition of a peroxidase-conjugated antibody to sheep (diluted 1:800) (Chemicon). Positive reactions were detected using 4-chloro-1-napthol.

Recombinant proteins were screened against pooled sera from five infected (strain VCS1001, serogroup A) sheep and five preimmune sheep by western blotting. Fractions containing comparable levels of protein were separated by SDS-PAGE, transferred to nitrocellulose membranes and incubated with sera (1:200) before the addition of a peroxidase-conjugated antibody to sheep (diluted 1:800) (Chemicon). Positive reactions were detected using 4-chloro-1-napthol.

Accession codes. GenBank: The complete annotated genome sequence is available at GenBank accession number CP000513. GEO: microarray data have been deposited at the GEO database under the series accession numbers GSE5157 (genomic microarrays) and GSE5166 (expression profiling).

Note: Supplementary information is available on the Nature Biotechnology website.

ACKNOWLEDGMENTS

The research was supported by an Initiative for Future Agriculture and Food Systems Grant No. 2001-52100 11445 from the USDA Cooperative

State Research, Education, and Extension Service, and by grants from the Australian Research Council. D.P was the recipient of an Australian Postgraduate Award and a Monash Faculty of Medicine, Nursing, and Health Sciences Postgraduate Excellence Award. We thank I. McPherson for technical assistance, C. Whitchurch for helpful discussions, B. Cheetham for providing unpublished information and helpful discussions, and the TIGR faculty, sequencing facility and informatics group for expert advice and assistance.

COMPETING INTERESTS STATEMENT

The authors declare no competing financial interests.

Published online at http://www.nature.com/naturebiotechnology
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions

1. Thomas, J.H. The pathogenesis of footrot in sheep with reference to proteases of *Fusiformis nodosus*. *Aust. J. Agric. Res.* **15**, 1001–1016 (1964).
2. Schwartzkoff, C.L. *et al.* The effects of antigenic competition on the efficacy of multivalent footrot vaccines. *Aust. Vet. J.* **70**, 123–126 (1993).
3. Stewart, D.J. Footrot of sheep. in *Footrot and Foot Abscess of Ruminants* (Egerton, J.R., Yong, W.K. & Riffkin, G.G., eds.) 5–45, (CRC Press, Boca Raton, Florida, USA, 1989).
4. Rood, J.I., Stewart, D.J., Vaughan, J.A. & Dewhirst, F.E. in *Bergey's Manual of Systematic Bacteriology Vol. 2: The Proteobacteria; Part B: The Gammaproteobacteria* Edn. 2 (Brenner, D.J., Kreig, N.R. & Staley, J.T., eds.) 124–129 (Springer, New York, 2005).
5. Calza, L., Manfredi, R. & Chiodo, F. Infective endocarditis: a review of the best treatment options. *Expert Opin. Pharmacother.* **5**, 1899–1916 (2004).
6. Kirkwood, J.K., Macgregor, S.K., Malnick, H. & Foster, G. Unusual mortality incidents in tit species (family Paridae) associated with the novel bacterium *Suttonella ornithocola*. *Vet. Rec.* **158**, 203–205 (2006).
7. Kennan, R.M., Billington, S.J. & Rood, J.I. Electroporation-mediated transformation of the ovine footrot pathogen *Dichelobacter nodosus*. *FEMS Microbiol. Lett.* **169**, 383–389 (1998).
8. Kennan, R.M., Dhungyel, O.P., Whittington, R.J., Egerton, J.R. & Rood, J.I. The Type IV fimbrial subunit gene (*fimA*) of *Dichelobacter nodosus* is essential for virulence, protease secretion, and natural competence. *J. Bacteriol.* **183**, 4451–4458 (2001).
9. Giovannoni, S.J. *et al.* Genome streamlining in a cosmopolitan oceanic bacterium. *Science* **309**, 1242–1245 (2005).
10. Moran, N.A. Microbial minimalism: genome reduction in bacterial pathogens. *Cell* **108**, 583–586 (2002).
11. Haring, V. *et al.* Delineation of the virulence-related locus (vrl) of *Dichelobacter nodosus*. *Microbiology* **141**, 2081–2089 (1995).
12. Billington, S.J., Johnston, J.L. & Rood, J.I. Virulence regions and virulence factors of the ovine footrot pathogen, *Dichelobacter nodosus*. *FEMS Microbiol. Lett.* **145**, 147–156 (1996).
13. Moses, E.K. *et al.* A multiple site-specific DNA-inversion model for the control of Omp1 phase and antigenic variation in *Dichelobacter nodosus*. *Mol. Microbiol.* **17**, 183–196 (1995).
14. Rood, J.I. Genomic islands of *Dichelobacter nodosus*. *Curr. Top. Microbiol. Immunol.* **264**, 47–60 (2002).
15. Cheetham, B.F. & Katz, M.E. A role for bacteriophages in the evolution and transfer of bacterial virulence determinants. *Mol. Microbiol.* **18**, 201–208 (1995).
16. Parker, D. *et al.* Regulation of type IV fimbrial biogenesis in *Dichelobacter nodosus*. *J. Bacteriol.* **188**, 4801–4811 (2006).
17. Claxton, P.D., Ribeiro, L.A. & Egerton, J.R. Classification of *Bacteroides nodosus* by agglutination tests. *Aust. Vet. J.* **60**, 331–334 (1983).
18. Hobbs, M. *et al.* Organization of the fimbrial gene region of *Bacteroides nodosus*: class I and class II strains. *Mol. Microbiol.* **5**, 543–560 (1991).
19. Sandkvist, M. Biology of type II secretion. *Mol. Microbiol.* **40**, 271–283 (2001).
20. Riffkin, M.C., Wang, L.F., Kortt, A.A. & Stewart, D.J. A single amino-acid change between the antigenically different extracellular serine proteases V2 and B2 from *Dichelobacter nodosus*. *Gene* **167**, 279–283 (1995).
21. Frey, J. & Kuhnert, P. RTX toxins in Pasteurellaceae. *Int. J. Med. Microbiol.* **292**, 149–158 (2002).
22. Suzuki, T. & Sasakawa, C. Molecular basis of the intracellular spreading of *Shigella*. *Infect. Immun.* **69**, 5959–5966 (2001).
23. Pizza, M. *et al.* Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing. *Science* **287**, 1816–1820 (2000).
24. Hess, J. *et al. Listeria monocytogenes* p60 supports host cell invasion by and in vivo survival of attenuated *Salmonella typhimurium*. *Infect. Immun.* **63**, 2047–2053 (1995).
25. Adler, B. *et al.* Candidate vaccine antigens and genes in *Pasteurella multocida*. *J. Biotechnol.* **73**, 83–90 (1999).
26. Yorgey, P., Rahme, L.G., Tan, M.W. & Ausubel, F.M. The roles of *mucD* and alginate in the virulence of *Pseudomonas aeruginosa* in plants, nematodes and mice. *Mol. Microbiol.* **41**, 1063–1076 (2001).
27. Konstantinidis, K.T. & Tiedje, J.M. Trends between gene content and genome size in prokaryotic species with larger genomes. *Proc. Natl. Acad. Sci. USA* **101**, 3160–3165 (2004).
28. Stover, C.K. *et al.* Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen. *Nature* **406**, 959–964 (2000).
29. Perez-Rueda, E. & Collado-Vides, J. The repertoire of DNA-binding transcriptional regulators in *Escherichia coli* K-12. *Nucleic Acids Res.* **28**, 1838–1847 (2000).
30. Egerton, J.R. & Burrell, D.H. Prophylactic and therapeutic vaccination against ovine foot-rot. *Aust. Vet. J.* **46**, 517–522 (1970).
31. Kennan, R.M., Dhungyel, O.P., Whittington, R.J., Egerton, J.R. & Rood, J.I. Transformation-mediated serogroup conversion of *Dichelobacter nodosus*. *Vet. Microbiol.* **92**, 169–178 (2003).
32. Zhou, H. & Hickford, J.G. Extensive diversity in New Zealand *Dichelobacter nodosus* strains from infected sheep and goats. *Vet. Microbiol.* **71**, 113–123 (2000).
33. Stewart, D.J., Clark, B.L., Emery, D.L., Peterson, J.E. & Fahey, K.J. A *Bacteroides nodosus* immunogen, distinct from the pilus, which induces cross-protective immunity in sheep vaccinated against footrot. *Aust. Vet. J.* **60**, 83–85 (1983).
34. Stewart, D.J. *et al.* The protection given by pilus and whole cell vaccines of *Bacteroides nodosus* strain 198 against ovine foot-rot induced by strains of different serogroups. *Aust. Vet. J.* **62**, 153–159 (1985).
35. Parker, D., Kennan, R.M., Myers, G.S., Paulsen, I. & Rood, J.I. Identification of a *Dichelobacter nodosus* ferric uptake regulator and determination of its regulatory targets. *J. Bacteriol.* **187**, 366–375 (2005).
36. Cianciotto, N.P. & Fields, B.S. *Legionella pneumophila mip* gene potentiates intracellular infection of protozoa and human macrophages. *Proc. Natl. Acad. Sci. USA* **89**, 5188–5191 (1992).
37. Leuzzi, R. *et al.* Ng-MIP, a surface-exposed lipoprotein of *Neisseria gonorrhoeae*, has a peptidyl-prolyl cis/trans isomerase (PPIase) activity and is involved in persistence in macrophages. *Mol. Microbiol.* **58**, 669–681 (2005).
38. Fraser, C.M. *et al.* Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*. *Nature* **390**, 580–586 (1997).
39. Tettelin, H., Radune, D., Kasif, S., Khouri, H. & Salzberg, S.L. Optimized multiplex PCR: efficiently closing a whole-genome shotgun sequencing project. *Genomics* **62**, 500–507 (1999).
40. Bulach, D.M. *et al.* Genome reduction in *Leptospira borgpetersenii* reflects limited transmission potential. *Proc. Natl. Acad. Sci. USA* **103**, 14560–14565 (2006).
41. Delcher, A.L., Phillippy, A., Carlton, J. & Salzberg, S.L. Fast algorithms for large-scale genome alignment and comparison. *Nucleic Acids Res.* **30**, 2478–2483 (2002).
42. Edgar, R.C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
43. Howe, K., Bateman, A. & Durbin, R. QuickTree: building huge Neighbour-Joining trees of protein sequences. *Bioinformatics* **18**, 1546–1547 (2002).
44. Thomas, J.H. A simple medium for the isolation and cultivation of *Fusiformis nodosus*. *Aust. Vet. J.* **34**, 411–413 (1958).
45. Gardy, J.L. *et al.* PSORT-B: Improving protein subcellular localization prediction for Gram-negative bacteria. *Nucleic Acids Res.* **31**, 3613–3617 (2003).
46. Nielsen, H., Engelbrecht, J., Brunak, S. & von Heijne, G. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.* **10**, 1–6 (1997).
47. Juncker, A.S. *et al.* Prediction of lipoprotein signal peptides in Gram-negative bacteria. *Protein Sci.* **12**, 1652–1662 (2003).
48. Cabrita, L.D., Dai, W. & Bottomley, S.P. A family of *E. coli* expression vectors for laboratory scale and high throughput soluble protein production. *BMC Biotechnol.* **6**, 12 (2006).