Author(s): Leo A. Calvo-Bado, Laura E. Green, Graham F. Medley, Atiya Ul-Hassan, Rose Grogono-Thomas, Nicky Buller, Jasmeet Kaler, Claire L. Russell, Ruth M. Kennan, Julian I. Rood and Elizabeth M.H. Wellington

Article Title: Detection and diversity of a putative novel heterogeneous polymorphic proline-glycine repeat (Pgr) protein in the footrot pathogen Dichelobacter nodosus

Year of publication: 2010

Link to published article: http://dx.doi.org/10.1016/j.vetmic.2010.06.024

Detection and diversity of a putative novel heterogeneous polymorphic proline-glycine repeat (Pgr) protein in the footrot pathogen *Dichelobacter nodosus*

Leo A. Calvo-Bado[a*], Laura E. Green[a], Graham F. Medley[a] , Atiya Ul-Hassan[a], Rose Grogono-Thomas[b], Nicky Buller[c], Jasmeet Kaler[a], Claire L. Russell[b], Ruth M. Kennan[d], Julian I. Rood[d], Elizabeth M.H. Wellington[a].

[a] *Department of Biological Sciences, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, UK.*

[b] *School of Clinical Veterinary Science, University of Bristol, Langford House, Langford BS40 5DU, UK.*

[c] *Animal Health Laboratories, Department of Agriculture and Food, Western Australia, 3 Baron-Hay Court, South Perth, WA6151, Australia.*

[d] *Australian Research Council Centre of Excellence in Structural and Functional Microbial Genomics, Department of Microbiology, Monash University, Victoria 3800, Australia.*

*Corresponding author: Tel.: +44(0)2476 522431; Fax: +44(0)2476 523701. *E-mail address*: l.a.calvo-bado@warwick.ac.uk (Leo A. Calvo-Bado).

Keywords: *Dichelobacter nodosus*, footrot, sheep, strain variation, repeat regions

Short Title: A novel proline-glycine repeat (Pgr) protein in *Dichelobacter nodosus*

ABSTRACT

Dichelobacter nodosus, a Gram negative anaerobic bacterium, is the essential causative agent of footrot in sheep. Currently, depending on the clinical presentation in the field, footrot is described as benign or virulent; *D. nodosus* strains have also been classified as benign or virulent, but this designation is not always consistent with clinical disease. The aim of this study was to determine the diversity of the *pgr* gene, which encodes a putative proline-glycine repeat protein (Pgr). The *pgr* gene was present in all 100 isolates of *D. nodosus* that were examined and, based on sequence analysis had two variants, *pgrA* and *pgrB*. In *pgrA*, there were two coding tandem repeat regions, R1 and R2: different strains had variable numbers of repeats within these regions. The R1 and R2 were absent from *pgrB*. Both variants were present in strains from Australia, Sweden and the UK, however, only *pgrB* was detected in isolates from Western Australia. The *pgrA* gene was detected in *D. nodosus* from tissue samples from two flocks in the UK with virulent footrot and only *pgrB* from a flock with no virulent or benign footrot for >10 years. Bioinformatic analysis of the putative PgrA protein indicated that it contained a collagen-like cell surface anchor motif. These results suggest that the *pgr* gene may be a useful molecular marker for epidemiological studies.

1. INTRODUCTION

The main causative agent of ovine footrot is *Dichelobacter nodosus,* a Gram-negative anaerobic bacterium with a small genome (Myers et al., 2007). The clinical manifestation of footrot is dependent on both environmental conditions (Depiazzi et al., 1998) and the virulence of the causative *D. nodosus* strain. Three laboratory tests have been developed to classify virulence of *D. nodosus* isolates. These assays are the protease thermostability or gelatin-gel test (Palmer, 1993), a test for the presence of the genetic element *intA* (Cheetham et al., 2006) and a test for the presence of the genomic islands *vap* and *vrl* (Rood et al., 1996). These tests do not always correlate, suggesting that either the tests do not test absolute markers for virulence or that virulence is complex and linked to more than one process.

From bioinformatic analysis of the VCS1703A genome of *D. nodosus* and comparative genome hybridization (CGH) of nine isolates (six virulent and three benign strains), a gene encoding a putative large, repetitive secreted protein, DNO_0690, was identified; designated here as Pgr for proline-glycine repeats. It was apparently present only in virulent strains and it was suggested that this gene might encode a virulence factor that could be involved in adhesion to the extracellular matrix (Myers et al., 2007).

The current study describes the isolation and diversity of DNO_0690 gene (*pgr*) in isolates from the UK, Australia and Sweden.

2. Materials and methods

2.1. Isolates

One hundred isolates from the UK (32), Australia (64) and Sweden (4) were analysed (Table I and Table II). The isolates included the nine Australian serogroup A-I prototype strains (VCS1001, VCS1006, VCS1008, VSC1172, VCS1137, VCS1017, VCS1220,

VCS1687, and VCS1623, respectively) and two virulent strains (VCS1703A and VCS1001 (also known as A198), one benign isolate (C305). There were 34 clinical isolates from Western Australia isolated from the feet of sheep at an abattoir (Buller et al., 2010). There were two isolates of bovine origin (C809 and C910) from the UK.

2.2. Growth and DNA extraction

Growth of *D. nodosus* and DNA extraction were done as described previously (Moore et al., 2005). DNA from the Australian isolates was extracted using Prepman Reagent (Applied Biosystems). DNA was extracted from the interdigital skin of the feet of three sheep each from two flocks with virulent footrot, and one flock with no virulent or benign footrot for >10 years. Tissue (130-160 mg) was treated with collagenase (10mg/ml) to release bacterial cells. Then the supernatant containing the extracted cells was pelleted at 15900 x g for 15 min. DNA was extracted using MagMAX™ Express Magnetic Particle Processors (AMBION, Applied Biosystems, Inc.) according to the manufacturer's recommendations.

2.3. PCR amplification, cloning and sequencing

The *pgr* gene from the sequenced genome of strain VCS1703A was used as a template to generate PCR primers. The *pgr* gene was detected in all isolates.

The complete *pgr* gene was amplified in 8 isolates (Fig. 1) with primers *pgrF* (5'-ATGGCAGTGATTACATTAAATGTTCGCTCCGCTAC-'3) and *pgrR* (5'-TTAGATGATTATGGAGCCAACGTGCCCATGTCAT-3') at an annealing temperature of 55^{o}C with KOD DNA polymerase master mix containing KOD Hot Start DNA polymerase (1.0 U/ µl), 10 × PCR Buffer for KOD, 25 mM $MgSO_4$, dNTP Mix (2 mM each), 1 × PCR buffer for KOD (Novagen). Amplicons were cloned into the pGEM-T Easy vector system (Promega) according to the manufacturer's recommendations. All amplicons were gel

purified (Qiagen) and sequenced using an ABI PRISM 3130xl Genetic Analyser (Applied Biosystems). Sequences were edited, aligned with SeqMan II (Lasergene 6) and MegAlign DNAstar software. Two variants of the *pgr* gene, *pgrA* and *pgrB*, were observed. These were differentiated by PCR with primers *pgrAF1* (5'- CCTGCACCATGCTTGTTAAA -'3) and *pgrAR1* (5'- GCTGTTGGTGGTTTGGCTAT -'3) at an annealing temperature of 60^{o}C and *pgrBF3* (5'-AKCATCRGGAAAGGTGA-'3) and *pgrBR2* (5'-GACGGCATCAGCAGCA-'3) at annealing temperature of 55^{o}C. The annealing temperatures were optimised for all PCR primers. Fifty isolates were tested against each set of primers for the *pgrA* and *pgrB* genes. Positive and negative controls were always included in each PCR reaction.

The R1 variable region was PCR-amplified and sequenced with primers *pgrF1* (5'-TTCCAACTGCTACCGTCA-'3) and *pgrR1* (5'-GCATTGGCAAGCGCAA-'3). All 100 sequences were submitted to the GenBank database (GU944975 to GU945040 and HM569229 to HM569262). All PCR amplifications were done using PCR-Promega master mix (Promega). Each sample contained 25 µl Master mix (50 units/ml of *Taq* DNA polymerase supplied in reaction buffer pH 8.5, 400 µM dATP, 400 µM dGTP, 400 µM dCTP, 400 µM dTTP, 3 mM MgCl2), 10 mM of each primer, 2.5 µl of DMSO (dimethyl sulfoxide, Fisher Scientific), 2 µl BSA (bovine serum albumin 10mg/ml, SIGMA) and 1-3 µl of template DNA (50-100 ng) in 50 µl reaction mix using the following conditions: 1 cycle of 95^{o}C for 2 min, 35 cycles of 95^{o}C for 1 min, 55^{o}C for 45 sec and 72^{o}C for 2 min with a final extension of 72^{o}C for 5min.

DNA from tissue from the UK flocks were screened initially by PCR for the presence of the *D. nodosus* specific 16S rRNA gene (*Cc* 5'-TCGGTACCGAGTATTTCTACCCAACACCT-'3 and *Ac* 5'-CGGGGTTATGTAGCTTGC-'3) (La Fontaine et al., 1993), then for presence of the *pgr* gene and then for the variants *pgrA* and *pgrB*.

2.4. Virulence tests

Protease thermostability (Palmer, 1993), the presence of *intA* (Cheetham et al., 2006) and the presence of genomic islands such as the *vap* and *vrl* regions (Katz et al., 1991; Rood et al., 1996) were carried out in the laboratory of origin on some samples as described in the relevant papers.

2.5. Phylogenetic analysis

For phylogenetic analyses, all nucleotide and protein sequences were aligned using ClustalW (www.ebi.ac.uk/ClustalW) and subsequently edited in Bioedit (www.mbio.ncsu.edu/Bioedit/bioedit.html). Neighbour joining trees were constructed using the Phylip package of programs (http://evolution.genetics.washington.edu/phylip.html). Distances between sequences were calculated in DNADIST and PROTDIST using the Kimura-2-parameter and the Jones-Taylor-Thornton model of sequence evolution for nucleotide and protein sequences respectively. Hypothetical basal PgrA and PgrB sequences were created *in silico* based on their sequence identity (all tandem repeats were removed); these sequences served as the root for tree constructions.

3. Results

3.1. Detection of the pgr gene in D. nodosus isolates and comparisons with other assays

All 100 isolates contained the *pgr* gene (Tables I and II). Sequence analysis revealed considerable variation, with two groupings of variants: *pgrA* (38/100 isolates) and *pgrB* (62/100 isolates). Of the isolates tested for protease activity, 28/31 (90%) *pgrA*-positive isolates had thermostable protease activity and 19/26 (73%) of *pgrB*-positive isolates had thermolabile protease activity (Table I). There was no correlation between *pgr* variant and

genetic elements such as *intA*, *vrl* and *vap* or serogroup. *PgrA* and *pgrB* were detected in isolates from the UK, Eastern Australia and Sweden. The UK isolates BS2, BS22, BS23, BS26 and BS27 that contained *pgrA* and isolates BS6, BS14, BS17, BS19, and BS24 that contained *pgrB* were isolated from feet with benign footrot (interdigital dermatitis), virulent footrot or Contagious Ovine Digital Dermatitis (CODD), suggesting that there was no correlation between the presence of *pgrA* or *pgrB* and the type of clinical presentation in the UK. All 34 isolates from WA had the *pgrB* variant; 59.5% (25) isolates had thermostable protease activity and 40.5% (17) had thermolabile protease activity (Table II).

3.2. Structure, phylogeny, diversity of the pgr gene and gene product

The *pgrA* gene in strain VSC1703A contained two coding regions with in-frame tandem repeats R1 and R2 at the 5' (18 nt) and 3' (54 nt) ends respectively. These varied in copy number in other strains. R1 was absent from the benign strain C305 (Fig. 1), which has the *prgB* variant. *PgrA* had 98-100% and *pgrB* <85% sequence identity to the *pgrA* gene from strain VCS1703A in the strains tested. Sequence variation in the *pgrA* derivatives was due to differences in the copy number of the 18 nt R1 region (GGTGAAAAACC(C/A)GGTAGT), which encoded a putative glycine rich repeat, GEKPGS. The number of repeats varied, for example there was one copy in isolate VCS1006 and six in VCS1703A; there was a maximum of 28 repeats in VCS1137 (Table I). The R2 region contained 54 nt (GACCCGGAAAATCCGACCAACCCGGTAGATCCTGAAAATCCAGATAAACCGAC T), which encoded a putative proline rich repeat, DPENPTNPVDPENPDKPT, which also varied in copy number from 1 to 16 repeats, these were in isolates BS1 and VCS1703A, respectively.

Isolates with the *pgrB* gene formed three sub-groups (Tables I and II); some isolates (AC390, T9015, UNE6, UNE13, 22477, 7004b and some WA isolates) had different tandem repeats that also varied in copy number; R3 (GGCGATGGAACTAAACCC; GDGTKP), R4 (AATCCTGGCGAAGGAACT; DDGTKP) or R5 (ATTTCCGGTATTACCCGTAGTACCTGAATCGCCGGTATTTCCTGTGTTTCCGGTA TTCCCCGTACTGCCTGAATCGCCGGT; TGDSGSTGNTGNTGNTGDSGTTGNTGN) while two isolates did not have any repeats (C395 and BS4). The *pgrB* region R5 was only observed in WA isolates, with 59.5% (25) isolates of these strains having this region (Table II).

Phylogenetic analysis of the sequences of the variable regions revealed that the isolates were grouped into four main clades that corresponded with the putative PgrA and PgrB proteins (Fig. 2). The PgrA clade had a long branch, and strains progressively increased the number of R1 repeats. The PgrB was more diverse, being subdivided into clades 1, 2 and 3. Clade 3 also had a long branch that reflected an increasing number of R3 repeats.

Bioinformatics analysis suggested that PgrA, but not PgrB, may be an outer membrane protein (http://www.expasy.org/tools/; http://bioinformatics.biol.uoa.gr) with two putative transmembrane helices and that the R1 and R2 regions represent collagen-like cell surface motifs (GEKPGS) and collagen alpha chain motifs (DPENPTNPVDPENPDKPT) respectively (http://www.ebi.ac.uk/Tools/InterProScan).

3.3. The pgr gene in samples and isolates from the UK and Western Australia.

All tissue samples from the three UK farms contained *D. nodosus* by 16S rRNA gene specific PCR. In sheep from the flock without footrot only the *pgrB* variant was detected, whilst in the other two flocks one or more *pgrA* variant with 7, 18 and 21 R1 repeats was detected. Finally, the two *D. nodosus* isolates from cattle were *pgrB* variants; both strains

were isolated from a 5 day old Belgian Blue cross from East Sussex, UK found dead with severe interdigital ulcerations on all four feet.

3.4. Design of a PCR assay to detect pgrA and pgrB variants

Isolates with the *pgrA* variant were discriminated from those with the *pgrB* variant by annealing temperatures of 60^{o}C and 55^{o}C respectively (Fig. 3). Subsequently, approximately 50 of the total 100 isolates were examined using this assay. The results correlated with the original *pgrA* and *pgrB* designations (Table I).

4. Discussion

In this study we have described the diversity of a gene encoding a putative novel and highly polymorphic proline-glycine repeat protein Pgr. This gene was present in all isolates of *D. nodosus* that we investigated. Sequence analysis of the *pgr* gene revealed that there were two variants both variants had regions with a variable number of tandem repeats. This variation, coupled with the fact that the Myers et al., (2007) results were derived from an oligonucleotide-based microarray might explain why Myers et al., (2007) thought that only virulent strains of *D. nodosus* contained Pgr.

Pgr is a putative protein with a molecular size of 100-120 kDa (881 to 1121 amino acid residues). The repeat regions in PgrA contain glycine rich (R1) and proline rich (R2) repeats and the hypervariability in copy number provides evidence that this protein is highly polymorphic. Proline-rich regions play an important role in many protein-protein interactions and cell-wall-spanning domains are preserved in many pathogenic bacteria (Williamson, 1994; Vanhoof et al., 1995; Kay et al., 2000; Girard et al., 2006). Collagen-like proteins are common in bacteria such as *Bartonella* (Zhang et al., 2004), *Streptococcus* (Lukomski et al., 2000; Paterson et al., 2008; Rasmussen et al., 2000) and *Yersinia* (Heise and Dersh, 2006)

and are involved in virulence, or mediating processes such as attachment, colonization and internalization.

Both the function of the Pgr products and the significance of the repeat regions are unknown. Variability in the number of in-frame tandem repeats has been associated with variation in virulence in other bacteria (Gravekamp et al., 1998; Nallapareddy et al., 2000; Puopolo et al., 2001; Jordan et al., 2003) and with human diseases (O'Dushlaine et al., 2005). These variations create protein polymorphisms and may alter antigenicity, as illustrated by the surface proteins alpha C of Group B streptococci (Gravekamp et al., 1996; Madoff et al., 1996) and Lmp1 and Lmp3 of *Mycoplasma hominis* (Ladefoged, 2000).

We postulate that the Pgr protein is at least partially exposed at the cell surface and that the observed polymorphisms might be a result of selective immune pressures in the host. The mechanisms that generate and maintain tandem repeats in bacterial genomes are poorly understood, but it has been suggested that they can arise from slipped strand mispairing, transformation and homologous recombination, or rolling circle replication (Romero and Palacios, 1997; Bzymek et al., 2001; Verstrepen et al., 2005). Since multiple strains of *D. nodosus* co-exist in individual feet and within flocks during subclinical and clinical infections (Claxton et al., 1983; Hindmarsh and Fraser, 1985; Jelinek et al., 2000; Moore et al., 2005), and antigenic changes can occur in *D. nodosus* following natural transformation and subsequent homologous recombination (Kennan et al., 2003), we suggest that both endogenous or exogenous *pgr* genes or gene regions may be responsible for the generation of these observed variants.

A footrot eradication programme has been in operation in WA for the last 20 years and the rate of isolation of gelatin-gel positive strains has declined with an increase in the prevalence of benign footrot. The 100% *pgrB* variant strains in sheep from WA suggests either that the eradication programme, which focussed on culling sheep with gelatin-gel

positive strains of *D. nodosus* may have excluded *pgrA* strains (this might be also be the case for the one UK flock with no signs of FR) alternatively, since previous data showed that *D. nodosus* isolates from WA were often different from other Australian isolates (Rood et al, 1996), it is possible that the *pgrA* gene might never have be present at detectible levels in WA. Cattle have been reported to harbour benign strains of *D. nodosus* (Pringel et al., 2008); the two cattle isolates in this study were *pgrB*, possibly indicating a species divide for these variants.

In summary, the *pgr* gene of *D. nodosus* is diverse but has two variants, *pgrA* and *pgrB*, which exhibit considerable heterogeneity. The *pgrA* variant has two repeat regions, generating a protein that may have collagen-like regions and cell surface anchor structures that are associated with invasion of the host in other pathogenic bacteria species. Discrimination of *pgr* variants with a PCR assay offers significant potential as a molecular tool in epidemiological studies.

Conflict of interest statement

None of the authors (L.A. Calvo-Bado, L.E Green, G.F. Medley, A. Ul-Hassan, R. Grogono-Thomas, N. Buller, J. Kaler, C.L. Russell, R.M. Kennan, J.I. Rood, and E.M.H. Wellington) has a financial or personal relationship with other people or organisation that could inappropriately influence or bias the paper entitled “Detection and diversity of a novel heterogeneous polymorphic proline-glycine repeat (Pgr) protein from the footrot pathogen *Dichelobacter nodosus*”.

Acknowledgments

This research was supported by the Biotechnology and Biological Sciences Research Council UK (BBRSC) under the Combating Endemic Diseases of Farmed Animals for

Sustainability (CEDFAS) initiative Grant No. BBE01870X1. RMK and JIR acknowledge the research support of the Australian Research Council through the ARC Centre of Excellence in Structural and Functional Microbial Genomics. We thank Brian Cheetham at the University of New England (Armidale, Australia), Viveca Baveraud at the National Veterinary Institute, Uppsala, Sweden, and Andrew Steventon at the Veterinary Laboratories Agency Winchester for some of the strains used in this study.

References

Buller N.B, Ashley P., Palmer M., Pitman D., Richards R.B., Hampson D.J., (2010). Understanding the molecular **e**pidemiology of the footrot pathogen *Dichelobacter nodosus* to support control and eradication programs. J. Clin. Microbiol. 48, 877-882.

Bzymek M., Lovett S.T., 2001. Instability of repetitive DNA sequences: The role of replication in multiple mechanisms. Proc. Natl. Acad. Sci. USA 98, 8319-8325.

Cheetham B.F., Tanjung L.R., Sutherland M., Druitt J., Green G., McFarlane J., Bailey G.D., Seaman J.T., Katz M.E., 2006. Improved diagnosis of virulent ovine footrot using the *intA* gene. Vet. Microbiol. 116, 166-174.

Claxton P.D., Ribeiro L.D., Egerton, J.R., 1983. Classification of *Bacteroides nodosus* by agglutination tests. Aust. Vet. J. 60, 331–334.

Depiazzi L.J., Roberts W.D., Hawkins C.D., Palmer M.A., Pitman D.R., Mcquade N.C., Jelinek P.D., Devereaux D.J., Rippon R.J., 1998. Severity and persistence of footrot in

299 Merino sheep experimentally infected with a protease thermostable strain of
300 *Dichelobacter nodosus* at five sites. Aust. Vet J. 76, 32-38.
301
302 Girard V., Mourez M., 2006. Adhesion mediated by autotransporters of Gram-negative
303 bacteria: Structural and functional features. Res. Microbiol. 157, 407-416.
304
305 Gravekamp C., Horensky D.S., Michel J.L., Madoff L.C., 1996. Variation in repeat number
306 within the alpha C protein of group B streptococci alters antigenicity and protective
307 epitopes. Infect. Immun. 64, 3576-3583.
308
309 Gravekamp C., Rosner B., Madoff L.C., 1998. Deletion of Repeats in the Alpha C Protein
310 Enhances the Pathogenicity of Group B Streptococci in Immune Mice. Infect. Immun. 66,
311 4347-4354.
312
313 Heise T., Dersch P., 2006. Identification of a domain in *Yersinia* virulence factor YadA that
314 is crucial for extracellular matrix-specific cell adhesion and uptake. PNAS 28, 3375–3380.
315
316 Hindmarsh F., Fraser J., 1985. Serogroups of *Bacteroides nodosus* isolated from ovine footrot
317 in Britain. Vet. Rec. 116, 187-188.
318
319 Jelinek P.D., Depiazzi L.J., Galvin D.A., Spicer I.T., Palmer M.A., Pitman D.R., 2000.
320 Occurrence of different strains of *Dichelobacter nodosus* in new clinical lesions in sheep
321 exposed to footrot associated with multi-strain infections. Aust. Vet. J. 78, 273-276.
322

Jordan P., Snyder L.A., Saunders N.J., 2003. Diversity in coding tandem repeats in related *Neisseria* spp. BMC Microbiol. 3, 23.

Katz M.E., Howarth P.M., Yong W.Y., Riffkin G.G., Depiazzi L.J., Rood J.I., 1991. Identification of three gene regions associated with virulence in *Dichelobacter nodosus*, the causative agent of ovine footrot, J. Gen. Microbiol. 137, 2117–2124.

Kay B., Williamson M.P., Sudol M., 2000. The importance of being proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains. FASEB 14, 231-241.

Kennan R.M., Dhungyel O.P., Whittington R.J., Egerton J.R., Rood J.I., 2003. Transformation-mediated serogroup conversion of *Dichelobacter nodosus*. Vet. Microbiol. 92, 169-78.

La Fontaine S., Egerton J.R., Rood J.I., 1993. Detection of *Dichelobacter nodosus* using species-specific oligonucleotides as PCR primers. Vet. Microbiol. 35, 101-117.

Ladefoged, S.A., 2000. Molecular dissection of *Mycoplasma hominis*. APMIS Suppl. 97, 1-45.

Lukomski S., Nakashima K., Abdi I., Cipriano V.J., Ireland R.M., Reid S.D., Adams G.G., Musser J.M., 2000. Identification and Characterization of the *scl* Gene Encoding a Group A *Streptococcus* Extracellular Protein Virulence Factor with Similarity to human Collagen. Infect. Immun. 68, 6542–6553.

348 Madoff L.C., Michel J.L., Gong E.W., Kling D.E., Kasper D.L., 1996. Group B streptococci
349 escape host immunity by deletion of tandem repeat elements of the alpha C protein. PNAS
350 93; 4131-4136.
351
352 Moore L.J., Wassink G.J., Green L.E., Grogono-Thomas R., 2005. The detection and
353 characterisation of *Dichelobacter nodosus* from cases of ovine footrot in England and
354 Wales. Vet. Microbiol. 108, 57-67.
355
356 Myers G.S., Parker D., Al-Hasani K., Kennan R.M., Seemann T., Ren Q., Badger J.H.,
357 Selengut J.D., Deboy R.T., Tettelin H., Boyce J.D., McCarl V.P., Han X., Nelson W.C.,
358 Madupu R., Mohamoud Y., Holley T., Fedorova N., Khouri H., Bottomley S.P.,
359 Whittington R.J., Adler B., Songer J.G., Rood J.I., Paulsen I.T., 2007. Genome sequence
360 and identification of candidate vaccine s from the animal pathogen *Dichelobacter*
361 *nodosus*. Nat. Biotechnol. 25, 569-575.
362
363 Nallapareddy S.R., Singh K.V., Duh R.W., Weinstock G.M., Murray B.E., 2000. Diversity of
364 ace, a gene encoding a microbial surface component recognizing adhesive matrix
365 molecules, from different strains of *Enterococcus faecalis* and evidence for production of
366 ace during human infections. Infect. Immun. 68, 5210-5217.
367
368 O'Dushlaine C.T., Edwards R.J., Park S.D., Shields D.C., 2005. Tandem repeat copy-number
369 variation in protein-coding regions of human genes. Genome Biol. 6, R69.
370
371 Palmer M.A., 1993. A gelatin test to detect activity and stability of proteases produced by
372 *Dichelobacter* (*bacteroides*) *nodosus*. Vet. Microbiol. 36, 113-122.

373 Paterson G.K., Nieminen L., Jefferies J.M.C., Mitchell T.J., 2008. PclA, a pneumococcal
374 collagen-like protein with selected strain distribution, contributes to adherence and
375 invasion of host cells. FEMS Microbiol. Lett. 285, 170–176.
376
377 Pringle M., Bergsten C., Fernström L.L., Höök H., Johansson K.E., 2008. Isolation and
378 characterization of *Treponema phagedenis*-like spirochetes from digital dermatitis lesions
379 in Swedish dairy cattle. Acta Vet. Scand. 50, 40-48.
380
381 Puopolo K.M., Hollingshead S.K., Carey V.J., Madoff L.C., 2001. Tandem repeat deletion in
382 the alpha C protein of group B *Streptococcus* is recA independent. Infect. Immun. 69,
383 5037-5045.
384
385 Rasmussen M., Edén A., Björck L., 2000. SclA, a novel collagen-like surface protein of
386 *Streptococcus pyogenes*. Infect. Immun. 68, 6370-6377.
387
388 Romero D., Palacios R., 1997. Gene amplification and genomic plasticity in prokaryotes.
389 Annu. Rev. Genet. 31, 91–111.
390
391 Rood J.I., Howarth P.A., Haring V., Billington S.J., Yong W.K., Liu D., Palmer M.A.,
392 Pitman D.R., Links I., Stewart D.J., Vaugham J.A., 1996. Comparison of gene probe and
393 conventional methods for the differentiation methods for the differentiation of ovine
394 footrot isolates of *Dichelobacter nodosus*. Vet. Microbiol. 52, 127-141.
395
396 Vanhoof G., Goossens F., Meester I.D., Hendriks D., Scharpe S., 1995. Proline motifs in
397 peptides and their biological processing. FASEB 9, 736-744.

Verstrepen K.J., Jansen A., Lewitter F., Fink G.R., 2005. Intragenic tandem repeats generate functional variability. Nat. Genet. 37, 986-990.

Williamson M.P., 1994. The structure and function of proline-rich regions in proteins. Biochem. J. 297, 249-260.

Zhang P., Chomel B.B., Schau M.K., Goo J.S., Droz S., Kelminson K.L., George S.S., Lerche N.W., Koehler J.E., 2004. A family of variably expressed outer-membrane proteins (Vomp) mediates adhesion and autoaggregation in *Bartonella quintana*. PNAS 101, 13630-13635.

Tables

Table I. Isolates, virulence parameters and source of strains

Table II. Isolates from Western Australia and virulence assays compared to *prg* test

Figure legends

Fig 1. Schematic representation of Pgr variants in *D. nodosus*. In the PgrA variants (VCS1006, VCS1623, VCS1220, VCS1703A and BS1) (A), the tandem repeat regions consist of 6 (R1, glycine repeats) and 18 (R2, proline repeats) amino acids. PgrB variants (B) have different tandem repeats of 6 amino acids for both regions (R3 and R4) (7004b). In some isolates the R3 and R4 regions are lacking (C305 and BS4).

Fig 2. Phylogenetic analysis of the deduced amino acid sequences of the putative Pgr protein from *D. nodosus*. The C-terminal region of the *pgr* genes of 100 isolates were analysed. The numbers on the branches indicate the percentage bootstrap value of 100 replicates and the scale bar indicates 10% nucleotide dissimilarity. Pgr sequences were grouped into in four main clades. PgrA isolates were in a single clade; PgrB were in three clades. A hypothetical PgrB sequence was included to serve as a root for tree constructions (see methods).

Fig 3. PCR assay specific for discrimination of *pgrA* and *pgrB* isolates. (a) *pgrA* primers, (b) *pgrB* primers. Lanes: MWM (molecular weight marker), BS1 (*pgrA*), BS2 (*pgrA*), BS3 (*pgrA*), BS4 (*pgrB*), BS5 (*pgrA*), BS6 (*pgrB*), BS11 (*pgrA*), BS12 (*pgrA*), BS14 (*pgrB*), water (Negative control), BS15 (*pgrA*), BS18 (*pgrA*), BS20 (*pgrA*), BS21 (*pgrB*), BS23 (*pgrA*), C305 (*pgrB*), C309 (*pgrB*), A198 (*pgrA*), C310 (*pgrB*).