

Nematology

Comparative phylogenetic analysis of vitellogenin in species of cyst and root-knot nematodes

--Manuscript Draft--

Manuscript Number:	NEMY-1659	
Full Title:	Comparative phylogenetic analysis of vitellogenin in species of cyst and root-knot nematodes	
Short Title:	Phylogenetic analysis of vitellogenin	
Article Type:	Full Research Paper	
Corresponding Author:	Hannah Wileman, BSc University of Hertfordshire Hatfield, Hertfordshire UNITED KINGDOM	
Keywords:	Caenorhabditis; development; egg; Globodera; Heterodera; Meloidogyne; sedentary endoparasite; vitellin	
Corresponding Author's Institution:	University of Hertfordshire	
First Author:	Hannah Wileman, BSc	
Order of Authors:	Hannah Wileman, BSc Keith G. Davies, PhD Roland N. Perry, PhD	
Manuscript Region of Origin:	UNITED KINGDOM	
Abstract:	<p>Summary</p> <p>Plant-parasitic nematodes (PPN) are an economically important group of crop pests and are oviparous animals; all nutrients required to develop and ensure the survival of their unhatched progeny need to be deposited within the egg, including proteins. The most abundant protein deposited is vitellin, formed of a precursor protein vitellogenin, which has roles in transporting lipids, providing amino acids and influencing post-embryonic development. The genes encoding vitellogenin have been well studied in <i>Caenorhabditis elegans</i>, but little is known about vitellogenin in PPN. Using the vitellogenin gene sequences from <i>C. elegans</i>, homologous sequences in the genomes of cyst and root-knot nematodes were identified and hypothetical vitellogenin genes were predicted. Protein domains were then determined. Sequences were aligned using MUSCLE and then used to construct phylogenetic trees using the maximum likelihood method. With the availability of genomic data and use of online local alignment tools, the vitellogenin encoding genes from <i>C. elegans</i> could be aligned to sequences from PPN genomes. All predicted genes contained the same protein domains as <i>C. elegans</i>; Vitellogenin_N, vitellogenin open beta-sheet and von Willebrand factor domain type D. The constructed phylogenetic tree clustered the species into three characterised groups, root-knot nematodes, cyst nematodes and <i>Caenorhabditis</i> species. Vitellogenin genes in <i>C. elegans</i> were homologous to sequences within PPN genomes, allowing the hypothetical genes to be determined and the relationships between PPN vitellogenin genes to be inferred, forming a potential basis to understand further the role of vitellogenin in cyst and root-knot nematodes.</p>	
Funding Information:	<p>K. G. Davies Ltd (HKEP University of Hertfordshire Project C004137)</p> <p>European Regional Development Fund (HKEP University of Hertfordshire Project C004137)</p> <p>Hertfordshire Local Enterprise Partnership (HKEP University of Hertfordshire Project C004137)</p>	<p>Miss Hannah Wileman</p> <p>Miss Hannah Wileman</p> <p>Miss Hannah Wileman</p>

1
2
3
4
5
6
7
8 **Comparative phylogenetic analysis of vitellogenin in species of cyst and**
9
10
11 **root-knot nematodes**
12
13
14
15

16
17 Hannah J. WILEMAN*, Roland N. PERRY and Keith G. DAVIES
18
19
20
21

22
23 *Department of Clinical, Pharmaceutical and Biological Sciences, School of Life and Medical*
24
25 *Sciences, University of Hertfordshire, Hatfield, Hertfordshire, AL10 9AB, UK*
26
27
28
29
30

31
32 Running title: Phylogenetic analysis of vitellogenin
33
34
35
36
37

38
39 *Corresponding author, e-mail: h.wileman3@herts.ac.uk
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Summary - Plant-parasitic nematodes (PPN) are an economically important group of crop pests and are oviparous animals; all nutrients required to develop and ensure the survival of their unhatched progeny need to be deposited within the egg, including proteins. The most abundant protein deposited is vitellin, formed of a precursor protein vitellogenin, which has roles in transporting lipids, providing amino acids and influencing post-embryonic development. The genes encoding vitellogenin have been well studied in *Caenorhabditis elegans*, but little is known about vitellogenin in PPN. Using the vitellogenin gene sequences from *C. elegans*, homologous sequences in the genomes of cyst and root-knot nematodes were identified and hypothetical vitellogenin genes were predicted. Protein domains were then determined. Sequences were aligned using MUSCLE and then used to construct phylogenetic trees using the maximum likelihood method. With the availability of genomic data and use of online local alignment tools, the vitellogenin encoding genes from *C. elegans* could be aligned to sequences from PPN genomes. All predicted genes contained the same protein domains as *C. elegans*; Vitellogenin_N, vitellogenin open beta-sheet and von Willebrand factor domain type D. The constructed phylogenetic tree clustered the species into three characterised groups, root-knot nematodes, cyst nematodes and *Caenorhabditis* species. Vitellogenin genes in *C. elegans* were homologous to sequences within PPN genomes, allowing the hypothetical genes to be determined and the relationships between PPN vitellogenin genes to be inferred, forming a potential basis to understand further the role of vitellogenin in cyst and root-knot nematodes.

Keywords - *Caenorhabditis briggsae*, *Caenorhabditis elegans*, development, egg, *Globodera*, *Heterodera*, *Meloidogyne*, plant-parasitic nematode, sedentary endoparasite, vitellin.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Plant-parasitic nematodes (PPN) have been estimated to cause at least \$US 80 billion of damage to crops per year (Nicol *et al.*, 2011). Infections by PPN cause non-specific symptoms such as stunted growth, wilting and yellowing of the leaves, caused by a decrease in function of the roots elicited by the nematode infection, which can lead to a reduced yield of the crop (Kumar & Yadav, 2020). Cyst nematodes (*Globodera* and *Heterodera*) and root-knot nematodes (*Meloidogyne*) are two of the most economically important groups of PPN (Jones *et al.*, 2013) and are found in the order Tylenchida.

Nematodes are oviparous animals; the developing nematode within the egg is surrounded by perivitelline fluid, which contains essential nutrients and protective antioxidants (Mkandawire *et al.*, 2021). For the successful development of the nematode, all nutrients for embryogenesis and further growth need to be deposited into the egg. Nutrients are deposited in the egg as proteins and other macromolecules; the primarily stored proteins are yolk proteins, also known as vitellin (Almenara *et al.*, 2013). Vitellin is formed of a precursor protein vitellogenin, which is a large glyco-lipoprotein complex (Winter *et al.*, 1996), a distant relative of the apoB protein in humans (Baker, 1988a). In *Caenorhabditis elegans*, synthesis of vitellogenin occurs in the hermaphrodite intestine, which is then secreted into the body cavity and taken up by the oocytes by receptor-mediated endocytosis (Perez & Lehner, 2019); the protein is then deposited in yolk granules of the egg (Winter *et al.*, 1996). Deposits of yolk protein remain within the egg after embryogenesis and prior to hatching (Bossinger & Schierenberg, 2003).

Vitellogenin is believed to have the primary function of the transport of lipids and micronutrients (Hayward *et al.*, 2010) and amino acid provision to developing progeny (Winter *et al.*, 1996). Additionally, vitellogenin has also been shown to influence post-embryonic phenotypes, whereby limited supply of the protein can result in smaller sized offspring and sterility in *C. elegans* (Perez & Lehner, 2019). Some studies report that

1 vitellogenin also serves a protective role against oxidative stress (Ishii *et al.*, 2002) and
2 environmental protection from bacterial infections (Fischer *et al.*, 2013).
3

4
5 In *C. elegans*, vitellogenin proteins consist of four polypeptides: two large, 170 kDa
6 yp170A and 170 kDa yp170B, and two small, 115 kDa yp115 and 88 kDa yp88 (Perez &
7 Lehner, 2019). Six genes encoding vitellogenin have been characterised in *C. elegans* (*vit* 1-
8 6) (Spieth *et al.*, 1985). The peptide encoded by *vit*-6 is proteolytically cleaved prior to
9 entering the oocyte, to form yp115 and yp88 (Spieth *et al.*, 1991).
10
11

12
13
14
15
16
17 Vitellogenin proteins and the associated encoding genes have also been characterised
18 in some species of both vertebrates (fish and birds) and invertebrates (insects).
19
20
21 *Caenorhabditis elegans* has been the primary model for investigating vitellogenin in
22 nematodes; however, orthologous genes have been found in other free-living nematodes:
23
24 *Oscheius tipulae* (Almenara *et al.*, 2013) and *C. briggsae* (Zucker-Aprison & Blumenthal,
25 1989). Vitellogenin has also been identified in the animal-parasitic nematodes *Toxocara*
26
27 *canis* (Zhu *et al.*, 2017), *Haemonchus contortus* (Hartman *et al.*, 2001) and *Trichostongylus*
28
29 *vitrinus* (Nisbet & Gasser, 2004). In *Heterodera glycines*, vitellogenin has been described and
30 partially characterised as two major egg proteins (Masler, 1999). These proteins constitute
31 over 50% of all proteins present within the egg with molecular weights of 190 kDa and 180
32
33 kDa, a comparative molecular weight to vitellogenin genes found in *C. elegans* (Masler,
34 1999). Additionally, when expressed sequence tag data from female *Globodera pallida*
35
36 samples were analysed, vitellogenin-like transcripts were found to be most abundant (Jones *et*
37
38 *al.*, 2009). Furthermore, the females of *G. pallida* develop from 21 to 35 days post infection
39
40 (dpi) of the host plant; vitellogenin was shown to be highly expressed at 28 and 35 dpi, due to
41
42 the deposition of yolk proteins within the oocytes (Cotton *et al.*, 2014). In the transcriptome
43
44 of the root-lesion nematode, *Pratylenchus penetrans*, vitellogenin was again shown to be
45
46 highly expressed (Vieira *et al.*, 2015).
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 Vitellogenin is an important protein in yolk provision for the developing nematode
2 and possibly has a protective role; however, they have not been identified or characterised in
3
4 many species of cyst and root-knot nematodes. With the increasing availability of genomic
5
6 data for many of these economically important species, it is now possible to determine
7
8 presence of vitellogenin genes within their genomes and infer their phylogenetic
9
10 relationships.
11
12

13 **Materials and methods**

14 15 16 17 18 19 20 21 22 DATA ACQUISITION

23
24
25
26 Vitellogenin gene sequences from *C. elegans* (*vit-1*, *vit-2*, *vit-3*, *vit-4*, *vit-5*, and *vit-6*)
27
28 were obtained using wormbase.org (Table S1). These sequences were then used as the query
29
30 sequences for BLAST searches against the genomes of selected species of cyst nematodes
31
32 (*Globodera* and *Heterodera* spp.), root-knot nematodes (*Meloidogyne* spp.) and *C. briggsae*
33
34 on the parasite.wormbase.org server. BLAST searches were completed using tblastn, with a
35
36 maximum of 100 target sequences, expect threshold of 0.01, and a word size of 2. The
37
38 scoring parameters used were BLOSUM62 matrix and low complexity regions were filtered.
39
40 The result gave homologous regions and a corresponding predicted ‘overlapping gene’ in
41
42 each genome, these gene sequences were then selected from the database and collated for
43
44 sequence alignment.
45
46
47
48
49

50 51 52 53 GENE PREDICTION

1 The genome of *G. ellingtonae* was not available from the server of
2 parasite.wormbase.org; instead NCBI BLAST was used to determine homologous regions
3 within the species genome. Regions of the genome containing homologous sequences from
4 the resulting BLAST searches were selected and uploaded into the Galaxy web platform
5 (Afgan *et al.*, 2018); here the sequences were then inputted into the AUGUSTUS tool (Stanke
6 & Morgenstern, 2005; Stanke *et al.*, 2008). *Caenorhabditis elegans* was chosen as the
7 training set model organism, the protein sequence and coding sequence were then predicted
8 using both strands. The predicted vitellogenin genes were then selected by comparing the
9 sequence to homologous regions identified in the BLAST search.
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25

26 PHYLOGENETIC TREE CONSTRUCTION

27
28
29
30
31 All genes, predicted or database derived through BLAST searches, were then
32 uploaded to Pfam (Bateman *et al.*, 2004) to determine theoretical protein domains for each
33 protein sequence. Sequences which contained the same protein domains as *C. elegans*
34 vitellogenin genes were included in the sequence alignment. Sequences identified to have
35 homologous regions to the original queries but did not exhibit all domains shown in *C.*
36 *elegans* vitellogenin were excluded from these data. Sequences were uploaded into Molecular
37 Evolutionary Genetics Analysis version ten software (MEGA X) (Kumar *et al.*, 2018) and
38 pairwise alignment was performed using the MUSCLE (multiple sequence comparison by
39 log-expectation) function in the software (Edgar, 2004). The cluster method used was
40 UPGMA.
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55

56 Phylogenetic trees were constructed using the maximum likelihood statistical method
57 in the MEGA X software, using the Jones-Taylor-Thornton substitution model, all sites
58
59
60
61
62
63
64
65

1 including gaps were used. For each tree constructed the test of phylogeny used was the
2 bootstrap method, this was replicated 100 times. Pairwise alignment scores were generated
3
4 using Jalview (Waterhouse *et al.*, 2009).
5
6
7
8

9 **Results**

10
11
12
13
14 The result from the BLAST search indicated that there were homologous regions and
15 genes to *C. elegans* vitellogenin within PPN genomes that have not previously been
16 identified. Two genes from *G. ellingtonae* were predicted and were shown to be closely
17 related to two *G. rostochiensis* vitellogenin genes (Fig. 1). Predicted vitellogenin gene *G.*
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

41 **PROTEIN DOMAINS IN VITELLOGENIN**

42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

1 present at the similar estimated positions. Similarly, all *Heterodera* vitellogenin genes and *G.*
2 *ellingtonae* 016 have protein domains in comparable positions to each other. Another set of
3 genes with domains in similar positions include all *G. rostochiensis* vitellogenin genes and *G.*
4 *pallida* 9459. Two *M. incognita* vitellogenin genes (28653 and 30865) have almost identical
5 domain positions. Finally, many of the *M. javanica* predicted vitellogenin genes are
6 comparable to each other excluding *M. javanica* 29487 and 43787.
7
8
9
10
11
12
13
14
15
16
17

18 PHYLOGENETIC RELATIONSHIPS OF VITELLOGENIN 19 20 21 22 23 24

25 A phylogenetic tree was constructed to infer the relationships between vitellogenin
26 genes in cyst nematodes, root-knot nematodes and Rhabditida species, *C. elegans* and *C.*
27 *briggsae* (Fig. 1). The tree was split into two major groups, the PPN species and the
28 *Caenorhabditis* species; within the PPN group this was further divided into the root-knot and
29 cyst nematode species with a small outgroup consisting of *G. pallida* 0564, *H. glycines* 13367
30 and *H. schachtii* 19700. The cyst nematodes species gave mixed locations for each gene and
31 were not clustered based on their species. As shown in Figure 1, there were three smaller
32 clusters of PPN each containing two, three and five more closely related genes. Two
33 predicted genes from *H. schachtii* genome were included in the dataset, *H. schachtii* 007 was
34 shown to be most closely related to *H. glycines* 23424, whereas *H. schachtii* 19700 was in the
35 outgroup of the cyst nematode cluster.
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50

51 The root-knot nematode cluster within Figure 1, comprising *M. hapla*, *M. incognita*
52 and *M. javanica*, was split into two main groups, with a small outgroup including the only *M.*
53 *hapla* hypothesised vitellogenin gene to possess all three protein domains. Like the cyst
54
55
56
57
58
59
60
61
62
63
64
65

1 nematodes, different genes from each species were in mixed locations and did not form
2 clusters of individual species.
3

4 The divergence in genes between *C. elegans* and a closely related species, *C.*
5 *briggsae*, was considered. Figure 1 gives *C. elegans vit-6* and *C. briggsae* 006 as an outgroup
6 of the *Caenorhabditis* cluster, where the remaining genes are split into two groups.
7 Vitellogenin genes from *C. elegans*, *vit-1*, *vit-2* and *vit-3*, were more closely related to *C.*
8 *briggsae* 002 and 16767, whereas *C. elegans vit-4* and *vit-5* were more closely related to *C.*
9 *briggsae* 14203 and 14234.
10
11
12
13
14
15
16
17
18
19
20
21
22

23 Discussion

24
25
26
27
28 This paper highlights the presence of vitellogenin genes within the genomes of cyst
29 and root-knot nematodes that have not been previously extensively studied. The presence of
30 vitellogenin genes in the genomes of PPN signifies a requirement for yolk provision to the
31 eggs of the developing nematode. A significant proportion of the vitellogenin amino acid
32 sequence is conserved between the cyst nematode, root-knot nematode and *Caenorhabditis*
33 vitellogenin genes; however, areas of divergence remain. Conservation in parts of the amino
34 acid sequence may be due to the functional domains of the proteins encoded, whereas the
35 divergence of the genes may be a result of speciation events from a common ancestor.
36
37
38
39
40
41
42
43
44
45
46
47

48 Multiple paralogous vitellogenin genes are present within many species of both
49 vertebrates and invertebrates, of differing amino acid composition (Smolenaars *et al.*, 2007).
50 In *C. elegans* these genes encode for different subunits of a larger protein complex, which
51 may allow for varying function of the encoded proteins (Perez & Lehner, 2019); it is possible
52 that multiple genes in PPN are required to form similar complexes. Data presented here
53
54
55
56
57
58
59
60
61
62
63
64
65

1 illustrate the relationship between these paralogous genes, with each species having differing
2 numbers of genes. Only one vitellogenin gene from the genome of *M. hapla* was identified
3 that contained all three protein domains, whereas the genome of *M. incognita* has seven
4 genes. This may be due to the relative sizes of their genomes, *M. hapla* genome contains 54
5 million base pairs (Opperman *et al.*, 2008), whilst the *M. incognita* genome contains 86
6 million base pairs (Abad *et al.*, 2008).
7
8
9
10
11
12
13

14 The genome of *G. ellingtonae* contains two not previously predicted vitellogenin
15 genes, which are both most closely related to two genes from *G. rostochiensis*. *Globodera*
16 *ellingtonae* has been shown to have similar behaviour and requirements for hatching and
17 development to *G. rostochiensis*, but shows some molecular diversity (Hesse *et al.*, 2021).
18 Like *G. rostochiensis*, *G. ellingtonae* is a pathogen of certain cultivars of potato but its
19 pathogenicity has been inconsistent (Zasada *et al.*, 2019). The inclusion of this recently
20 discovered potato cyst nematode, *G. ellingtonae*, sought to further evaluate the nematode
21 species against other members of the *Globodera* family.
22
23
24
25
26
27
28
29
30
31
32
33

34 As vitellogenin plays a key role in the development of the nematode, it could be
35 anticipated that it would have a relatively conserved sequence, especially between closely
36 related species. This has also been illustrated in the comparison of *C. elegans* with *C.*
37 *briggsae* where sequences were shown to be 85% identical in coding regions (Zucker-
38 Aprison & Blumenthal, 1989); therefore, vitellogenin genes in *C. elegans* and *C. briggsae* are
39 highly conserved. The genome of *C. elegans* contains six vitellogenin genes whereas *C.*
40 *briggsae* only contains five (Zucker-Aprison & Blumenthal, 1989). Perez & Lehner (2019)
41 suggest that the *vit-4* gene is missing from *C. briggsae* due to a duplication between *vit-3* and
42 *vit-4*. Contrary to this, data presented here indicate that there are two genes closely related to
43 *C. elegans vit-4* and *vit-5* in *C. briggsae*, but only two genes closely related to *C. elegans vit-*
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1, *vit-2* and *vit-3*. This result suggests that either *vit-1*, *vit-2* or *vit-3* may be missing from *C. briggsae*, not *vit-4* as originally predicted.

Vitellogenin genes have been shown to be upregulated in *G. pallida*. In supplementary data provided by Cotton *et al.* (2014), three *G. pallida* vitellogenin genes were shown to be upregulated in females (at 21, 28 and 35 dpi) compared to the early parasitic life-stage (7 and 14 dpi). The three genes shown to be upregulated are presented here and form part of the main cyst nematode cluster in the phylogenetic analyses. The gene *G. pallida* 0564, which forms part of an outgroup from the PPN cluster, is not shown to be expressed in that study; it is possible that this gene was not identified, or this gene is not expressed. Currently, no vitellogenin genes have been characterised in *G. rostochiensis*.

The genomes of species within the cyst nematode genus *Heterodera* contain vitellogenin genes; this study highlights their presence within the genomes of *H. schachtii* and *H. glycines*. Two *H. schachtii* genes are shown to be closely related to those in *H. glycines*; however, a third *H. glycines* gene appears to be more divergent. This result supports the findings of Singh *et al.* (2020), who generated a phylogenetic consensus tree containing 33 different populations of *Heterodera* spp., finding that *H. schachtii* and *H. glycines* were closely related. In a study that sequences the *H. schachtii* genome, data provided in the supplementary information listed two proteins that were annotated as containing a lipoprotein amino terminal region, von Willebrand factor type D domain and a domain of unknown function, whilst also indicating that they were highly expressed at 24 dpi (Siddique *et al.*, 2021). These two proteins have been incorporated into Figure 1 as HS_19700 and HS_007; establishing that they also have a vitellogenin open beta-sheet domain, which was not previously annotated.

Masler (1999) partially characterised two vitellogenin proteins in *H. glycines* with similar molecular weights (180 and 190 kDa) to the two larger protein subunits of *C. elegans*

1 (170 kDa yp170A and 170 kDa yp170B). Figure 1 presents three *H. glycines* hypothesised
2 vitellogenin genes and it is possible that these genes encode for the ‘major egg proteins’ as
3 described (Masler, 1999). The *H. glycines* transcriptome of effectors gave only one predicted
4 vitellogenin gene sequence, provided in the supplementary information (Gardner *et al.*,
5 2018). Again, it may be possible that not all vitellogenin genes present in the genome are
6 expressed to produce the vitellogenin proteins.
7

8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Root-knot nematodes are the most damaging nematode crop pests, so it was important to include some species of this genus. The relationship between 13 vitellogenin genes from root-knot nematode species has been illustrated, including one gene from the genome of *M. hapla*, five from *M. javanica* and seven from the genome of *M. incognita*. Currently, no studies have highlighted the presence of vitellogenin within the genomes of *M. hapla* or *M. javanica*. In previous analysis of the *M. incognita* genome, no vitellogenin encoding genes were identified (Abad *et al.*, 2008); however, in supplementary data provided by Bellafiore *et al.* (2008), five vitellogenin genes were found in the secreted effectors of female pharyngeal glands using mass spectrometry. A similar study of the effectors produced by female pharyngeal glands in *G. rostochiensis* resulted in no vitellogenin genes being detected (Maier *et al.*, 2012). The result from Bellafiore *et al.* (2008) is unexpected, as it is understood from *C. elegans* that vitellogenin genes are primarily expressed in the hermaphrodite intestine and then transported into the oocyte by receptor-mediated endocytosis (Perez & Lehner, 2019). However, it has now been shown that the *C. elegans* hermaphrodite vents vitellogenin as part of the yolk after egg laying ceases, for further nutritional provision for the developing nematode (Kern *et al.*, 2021). It may be possible that root-knot nematodes exhibit a similar behaviour when producing an egg mass, which primarily consists of glycoproteins (Sharon & Spiegel, 1993).

1 Protein domains present in *C. elegans* vitellogenin appear within the identified
2 vitellogenin genes of cyst and root-knot nematodes. Using Pfam to predict the location of
3
4 protein domains, *C. elegans*, *C. briggsae*, cyst and root-knot nematode vitellogenin genes
5
6 were shown to contain three protein domains: Vitellogenin_N, vitellogenin open beta-sheet
7
8 and von Willebrand factor type D domain. In BLAST searches of PPN genomes, some genes
9
10 contained only one or two of these domains, with a sequence length that was much shorter
11
12 than other vitellogenin genes within the species; therefore, these results were not included in
13
14 the analysis. Protein domain Vitellogenin_N is a lipoprotein amino terminal region that is
15
16 predicted to be involved in lipid transport and is a domain present in the vitellogenin of
17
18 certain insect, crustacean and nematode species (Smolenaars *et al.*, 2007). Nematode
19
20 vitellogenin have been shown to contain amphipathic β -strands, which may also function in
21
22 the transport of additional lipids (Smolenaars *et al.*, 2007). These strands may then form the
23
24 vitellogenin open beta-sheet domain. Towards the C-terminus of the vitellogenin gene in
25
26 nematodes, the protein domain von Willebrand type D is present. This domain has shown
27
28 adhesive properties and has been hypothesised to be the binding site for membrane receptors
29
30 on the oocyte (Baker, 1988b).
31
32
33
34
35
36
37
38

39 Whilst the presence of vitellogenin genes within certain PPN species has been
40
41 determined, a complete understanding of their expression and function is yet to be fully
42
43 elucidated. Each nematode species contains a different number of vitellogenin genes, even
44
45 between closely related species. There is a possibility that the variance in the number of
46
47 genes may be due to differences in expression, with some genes not being expressed. Further
48
49 analysis of the stage- and tissue-specific expression of vitellogenin would help to understand
50
51 further their presence and role in PPN. Vitellogenin genes in selected cyst and root-knot
52
53 nematode species contain the same protein domains as those in *C. elegans* so could be
54
55 hypothesised to exhibit a similar function; however, this requires further investigation.
56
57
58
59
60
61
62
63
64
65

1 Masler (1999) has partially characterised vitellogenin proteins in *H. glycines*; this work could
2 be expanded to characterise the proteins in other PPN species. Vitellogenin plays a key role
3 in the transport of lipids and nutritional provision to the developing nematode; if this function
4 could be interrupted this may provide a novel route for control of PPN. Additionally,
5 vitellogenin has been shown to have a role in protection against environmental stress in *C.*
6
7
8
9
10
11
12 *elegans* (Fischer *et al.*, 2013). This has been demonstrated by the knockdown of *vit-6* gene by
13
14 RNA-interference, which increased the nematodes susceptibility to *Photorhabdus*
15
16 *luminescens*, reducing its lifespan (Fischer *et al.*, 2013). If vitellogenin inhibited in PPN, this
17
18 may provide a novel method of control by potentially increasing the nematode's
19
20
21
22
23 susceptibility to environmental stress.

24 This study set out to highlight the presence of and predict vitellogenin genes within
25
26 the genomes of some economically important PPN. The genes encoding the vitellogenin
27
28 protein in *C. elegans* were found to be homologous to sequences in the genomes of cyst and
29
30 root-knot nematodes and could be used to infer relationships between the PPN species.
31
32
33 Additionally, the protein domains of vitellogenin genes that appear in *C. elegans* were also
34
35 present in the PPN. This study contributes to understanding of vitellogenin in PPN; however,
36
37
38 further investigation needs to be completed to fully characterise the proteins and understand
39
40
41 their biological role. By determining the presence of vitellogenin genes within the genomes
42
43
44 of some economically important PPN, the relationship between these genes could be inferred
45
46
47 and this information could help form a basis for further investigation of vitellogenin in PPN.

51 **Acknowledgements**

52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 This research was supported by funding from 1) Norwegian Institute of Bioeconomy
2 Research, 2) K. G. Davies Ltd, 3) European Regional Development Fund, 4) Hertfordshire
3 Local Enterprise Partnership.
4
5
6
7
8
9

10 **References**

- 11
12
13
14
15 ABAD, P., GOUZY, J., AURY, J.M., CASTAGNONE-SERENO, P., DANCHIN, E.G., DELEURY, E.,
16
17 PERFUS-BARBEOCH, L., ANTHOUARD, V., ARTIGUENAVE, F., BLOK, V.C. *et al.* (2008).
18
19 Genome sequence of the metazoan plant-parasitic nematode *Meloidogyne*
20
21 *incognita*. *Nature Biotechnology* 26, pp. 909-915. DOI: 10.1038/nbt.1482
22
23
24
25 AFGAN, E., BAKER, D., BATUT, B., VAN DEN BEEK, M., BOUVIER, D., ČECH, M., CHILTON, J.,
26
27 CLEMENTS, D., CORAOR, N., GRÜNING, B.A. *et al.* (2018). The Galaxy platform for
28
29 accessible, reproducible, and collaborative biomedical analyses: 2018 update. *Nucleic*
30
31 *Acids Research* 46, pp. 537-544. DOI:10.1093/nar/gky379
32
33
34
35 ALMENARA, D.P., DE MOURA, J.P., SCARABOTTO, C.P., ZINGALI, R.B. & WINTER, C.E. (2013).
36
37 The molecular and structural characterization of two vitellogenins from the free-living
38
39 nematode *Oscheius tipulae*. *PloS One* 8, e53460. DOI: 10.1371/journal.pone.0053460
40
41
42 BAKER, M.E., (1988a). Is vitellogenin an ancestor of apolipoprotein B-100 of human low-
43
44 density lipoprotein and human lipoprotein lipase?. *Biochemical Journal* 255, pp.
45
46 1057-1060. DOI: 10.1042/bj2551057.
47
48
49 BAKER, M.E., (1988b). Invertebrate vitellogenin is homologous to human von Willebrand
50
51 factor. *Biochemical Journal* 256, p. 1059. DOI: 10.1042/bj2561059
52
53
54 BATEMAN, A., COIN, L., DURBIN, R., FINN, R.D., HOLLICH, V., GRIFFITHS-JONES, S., KHANNA,
55
56 A., MARSHALL, M., MOXON, S., SONNHAMMER, E.L. *et al.* (2004). The Pfam protein
57
58 families database. *Nucleic Acids Research* 32, pp. 138-141. DOI: 10.1093/nar/gkh121
59
60
61
62
63
64
65

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
- BELLAFIGLIORE, S., SHEN, Z., ROSSO, M.N., ABAD, P., SHIH, P. AND BRIGGS, S.P. (2008). Direct identification of the *Meloidogyne incognita* secretome reveals proteins with host cell reprogramming potential. *PLoS Pathogens* 4, e1000192. DOI: 10.1371/journal.ppat.1000192
- BOSSINGER, O. & SCHIERENBERG, E. (2003). The use of fluorescent marker dyes for studying intercellular communication in nematode embryos. *International Journal of Developmental Biology* 40, pp. 431-439.
- COTTON, J.A., LILLEY, C.J., JONES, L.M., KIKUCHI, T., REID, A.J., THORPE, P., TSAI, I.J., BEASLEY, H., BLOK, V., COCK, P.J. *ET AL.* (2014). The genome and life-stage specific transcriptomes of *Globodera pallida* elucidate key aspects of plant parasitism by a cyst nematode. *Genome Biology* 15, pp. 1-17. DOI: 10.1186/gb-2014-15-3-r43
- EDGAR, R.C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5, pp. 1-19. DOI: 10.1186/1471-2105-5-113
- FISCHER, M., REGITZ, C., KULL, R., BOLL, M. & WENZEL, U. (2013). Vitellogenins increase stress resistance of *Caenorhabditis elegans* after *Photorhabdus luminescens* infection depending on the steroid-signaling pathway. *Microbes and Infection* 15, pp. 569-578. DOI: 10.1016/j.micinf.2013.05.002
- GARDNER, M., DHROSO, A., JOHNSON, N., DAVIS, E.L., BAUM, T.J., KORKIN, D. & MITCHUM, M.G. (2018). Novel global effector mining from the transcriptome of early life stages of the soybean cyst nematode *Heterodera glycines*. *Scientific Reports* 8, pp. 1-15. DOI: 10.1038/s41598-018-20536-5
- HAYWARD, A., TAKAHASHI, T., BENDENA, W.G., TOBE, S.S. & HUI, J.H. (2010). Comparative genomic and phylogenetic analysis of vitellogenin and other large lipid transfer

1 proteins in metazoans. *FEBS Letters* 584, pp. 1273-1278. DOI:
2 10.1016/j.febslet.2010.02.056
3

4 HESSE, C.N., MORENO, I., PARDO, O.A., FUENTES, H.P., GRENIER, E., DANDURAND, L.M. &
5 ZASADA, I.A. (2021). Characterization of Populations from Chile Utilizing Whole
6 Genome Sequencing. *Journal of Nematology* 53, pp. 1-9. DOI: 10.21307/jofnem-
7 2021-088
8

9 ISHII, N., GOTO, S. & HARTMAN, P.S. (2002). Protein oxidation during aging of the nematode
10 *Caenorhabditis elegans*. *Free Radical Biology and Medicine* 33, pp. 1021-1025. DOI:
11 10.1016/S0891-5849(02)00857-2
12

13 JONES, J.T., KUMAR, A., PYLYPENKO, L.A., THIRUGNANASAMBANDAM, A., CASTELLI, L.,
14 CHAPMAN, S., COCK, P.J., GRENIER, E., LILLEY, C.J., PHILLIPS, M.S. *et al.* (2009).
15 Identification and functional characterization of effectors in expressed sequence tags
16 from various life cycle stages of the potato cyst nematode *Globodera*
17 *pallida*. *Molecular Plant Pathology* 10, pp. 815-828. DOI: 10.1111/j.1364-
18 3703.2009.00585.x
19

20 JONES, J.T., HAEGEMAN, A., DANCHIN, E.G., GAUR, H.S., HELDER, J., JONES, M.G., KIKUCHI,
21 T., MANZANILLA-LÓPEZ, R., PALOMARES-RIUS, J.E., WESEMAEL, W.M. *et al.* (2013).
22 Top 10 plant-parasitic nematodes in molecular plant pathology. *Molecular Plant*
23 *Pathology* 14, pp. 946-961. DOI: 10.1111/mpp.12057
24

25 KERN, C.C., TOWNSEND, S., SALZMANN, A., RENDELL, N.B., TAYLOR, G.W., COMISEL, R.M.,
26 FOUKAS, L.C., BÄHLER, J. & GEMS, D. (2021). *C. elegans* feed yolk to their young in a
27 form of primitive lactation. *Nature Communications* 12, pp. 1-11. DOI:
28 10.1038/s41467-021-25821-y
29
30
31
32
33
34
35

- 1 KUMAR, S., STECHER, G., LI, M., KNYAZ, C., & TAMURA, K. (2018). MEGAX: molecular
2 evolutionary genetics analysis across computing platforms. *Molecular Biology and*
3 *Evolution* 35, p. 1547. DOI: 10.1093/molbev/msy096
4
5
6
7 KUMAR, Y. & YADAV, B.C. (2020). Plant-parasitic nematodes: Nature's most successful plant
8 parasite. *International Journal of Research and Review* 7, pp. 379-386.
9
10
11
12 MAIER, T.R., HEWEZI, T., PENG, J. & BAUM, T.J. (2013). Isolation of whole esophageal gland
13 cells from plant-parasitic nematodes for transcriptome analyses and effector
14 identification. *Molecular Plant-Microbe Interactions* 26, pp. 31-35. DOI:
15 10.1094/MPMI-05-12-0121-FI
16
17
18
19
20
21
22 MASLER, E.P. (1999). Detection and partial characterization of egg polypeptides from
23 *Heterodera glycines*. *Journal of Nematology* 31, p. 305-311.
24
25
26
27 MKANDAWIRE, T.T., GRENCIS, R.K., BERRIMAN, M. & DUQUE-CORREA, M.A. (2021).
28 Hatching of parasitic nematode eggs: a crucial step determining infection. *Trends in*
29 *Parasitology* 38, pp. 174-187. DOI: 10.1016/j.pt.2021.08.008
30
31
32
33
34 NICOL, J.M., TURNER, S.J., COYNE, D.L., NIJS, L.D., HOCKLAND, S. & MAAFI, Z.T. (2011).
35 Current nematode threats to world agriculture. In: Jones, J., Gheyson, G. & Fenoll, C.
36 (Eds). *Genomics and Molecular Genetics of Plant-Nematode Interactions*. Dordrecht,
37 Netherlands, Springer, pp. 21-43. DOI: 10.1007/978-94-007-0434-3_2
38
39
40
41
42
43
44 OPPERMAN, C.H., BIRD, D.M., WILLIAMSON, V.M., ROKHSAR, D.S., BURKE, M., COHN, J.,
45 CROMER, J., DIENER, S., GAJAN, J., GRAHAM, S. *et al.* (2008). Sequence and genetic
46 map of *Meloidogyne hapla*: A compact nematode genome for plant
47 parasitism. *Proceedings of the National Academy of Sciences* 105, pp. 14802-14807.
48
49
50
51
52
53
54 DOI: 10.1073/pnas.0805946105
55
56
57
58
59
60
61
62
63
64
65

- 1 PEREZ, M.F. & LEHNER, B. (2019). Vitellogenins- yolk gene function and regulation in
2 *Caenorhabditis elegans*. *Frontiers in Physiology* 10, p. 1067. DOI:
3
4 10.3389/fphys.2019.01067
5
6
7 SHARON, E. & SPIEGEL, Y. (1993). Glycoprotein characterization of the gelatinous matrix in
8
9 the root-knot nematode *Meloidogyne javanica*. *Journal of Nematology* 25, p.585-589.
10
11
12 SIDDIQUE, S., RADAKOVIC, Z.S., HILTL, C., PELLEGRIN, C., BAUM, T.J., BEASLEY, H.,
13
14 CHITAMBO, O., CHOPRA, D., DANCHIN, E.G., GRENIER, E. *et al.* (2021). The genome
15
16 and life stage-specific transcriptomes of a plant-parasitic nematode and its host reveal
17
18 susceptibility genes involved in trans-kingdom synthesis of vitamin B. *bioRxiv*. DOI:
19
20 10.1101/2021.10.01.462558
21
22
23
24 SINGH, P.R., KARSSSEN, G., COUVREUR, M. & BERT, W. (2020). Morphological and molecular
25
26 characterization of n. sp. (Nematoda: Heteroderidae) from Gran Canaria, Canary
27
28 Islands. *Journal of Nematology* 52, pp. 1-14. DOI: 10.21307/jofnem-2020-098
29
30
31
32 SMOLENAARS, M.M., MADSEN, O., RODENBURG, K.W. & VAN DER HORST, D.J. (2007).
33
34 Molecular diversity and evolution of the large lipid transfer protein
35
36 superfamily. *Journal of Lipid Research* 48, pp. 489-502. DOI: 10.1194/jlr.R600028-
37
38 JLR200
39
40
41 SPIETH, J., DENISON, K., ZUCKER, E. & BLUMENTHAL, T., (1985). The nucleotide sequence of
42
43 a nematode vitellogenin gene. *Nucleic Acids Research* 13, pp. 7129-7138. DOI:
44
45 10.1093/nar/13.19.7129
46
47
48
49 SPIETH, J., NETTLETON, M., ZUCKER-APRISON, E., LEA, K. & BLUMENTHAL, T. (1991).
50
51 Vitellogenin motifs conserved in nematodes and vertebrates. *Journal of Molecular*
52
53 *Evolution* 32, pp. 429-438. DOI: 10.1007/BF02101283
54
55
56
57
58
59
60
61
62
63
64
65

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
- STANKE, M., DIEKHANS, M., BAERTSCH, R. & HAUSSLER, D. (2008). Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* 24, pp. 637-644. DOI: 10.1093/bioinformatics/btn013
- STANKE, M. & MORGENSTERN, B. (2005). AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Research* 33, pp. 465-467. DOI: 10.1093/nar/gki458
- VIEIRA, P., EVES-VAN DEN AKKER, S., VERMA, R., WANTOCH, S., EISENBACK, J.D. & KAMO, K. (2015). The *Pratylenchus penetrans* transcriptome as a source for the development of alternative control strategies: mining for putative genes involved in parasitism and evaluation of in planta RNAi. *PLoS One* 10, e0144674. DOI: 10.1371/journal.pone.0144674
- WATERHOUSE, A.M., PROCTER, J.B., MARTIN, D.M., CLAMP, M. & BARTON, G.J. (2009). Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25, pp. 1189-1191. DOI: 10.1093/bioinformatics/btp033
- WINTER, C.E., PENHA, C. & BLUMENTHAL, T. (1996). Comparison of a vitellogenin gene between two distantly related rhabditid nematode species. *Molecular Biology and Evolution* 13, pp. 674-684. DOI: 10.1093/oxfordjournals.molbev.a025628
- ZASADA, I. A., INGHAM, R. E., BAKER, H., & PHILLIPS, W. S. (2019). Impact of *Globodera ellingtonae* on yield of potato (*Solanum tuberosum*). *Journal of Nematology* 51. DOI: 10.21307/JOFNEM-2019-073
- ZUCKER-APRISON, E. & BLUMENTHAL, T. (1989). Potential regulatory elements of nematode vitellogenin genes revealed by interspecies sequence comparison. *Journal of Molecular Evolution* 28, pp. 487-496. DOI: 10.1007/BF02602929

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

**Comparative phylogenetic analysis of vitellogenin in species of cyst and
root-knot nematodes**

Running title: Phylogenetic analysis of vitellogenin

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Summary - Plant-parasitic nematodes (PPN) are an economically important group of crop pests and are oviparous animals; all nutrients required to develop and ensure the survival of their unhatched progeny need to be deposited within the egg, including proteins. The most abundant protein deposited is vitellin, formed of a precursor protein vitellogenin, which has roles in transporting lipids, providing amino acids and influencing post-embryonic development. The genes encoding vitellogenin have been well studied in *Caenorhabditis elegans*, but little is known about vitellogenin in PPN. Using the vitellogenin gene sequences from *C. elegans*, homologous sequences in the genomes of cyst and root-knot nematodes were identified and hypothetical vitellogenin genes were predicted. Protein domains were then determined. Sequences were aligned using MUSCLE and then used to construct phylogenetic trees using the maximum likelihood method. With the availability of genomic data and use of online local alignment tools, the vitellogenin encoding genes from *C. elegans* could be aligned to sequences from PPN genomes. All predicted genes contained the same protein domains as *C. elegans*; Vitellogenin_N, vitellogenin open beta-sheet and von Willebrand factor domain type D. The constructed phylogenetic tree clustered the species into three characterised groups, root-knot nematodes, cyst nematodes and *Caenorhabditis* species. Vitellogenin genes in *C. elegans* were homologous to sequences within PPN genomes, allowing the hypothetical genes to be determined and the relationships between PPN vitellogenin genes to be inferred, forming a potential basis to understand further the role of vitellogenin in cyst and root-knot nematodes.

Keywords - *Caenorhabditis briggsae*, *Caenorhabditis elegans*, development, egg, *Globodera*, *Heterodera*, *Meloidogyne*, plant-parasitic nematode, sedentary endoparasite, vitellin.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Plant-parasitic nematodes (PPN) have been estimated to cause at least \$US 80 billion of damage to crops per year (Nicol *et al.*, 2011). Infections by PPN cause non-specific symptoms such as stunted growth, wilting and yellowing of the leaves, caused by a decrease in function of the roots elicited by the nematode infection, which can lead to a reduced yield of the crop (Kumar & Yadav, 2020). Cyst nematodes (*Globodera* and *Heterodera*) and root-knot nematodes (*Meloidogyne*) are two of the most economically important groups of PPN (Jones *et al.*, 2013) and are found in the order Tylenchida.

Nematodes are oviparous animals; the developing nematode within the egg is surrounded by perivitelline fluid, which contains essential nutrients and protective antioxidants (Mkandawire *et al.*, 2021). For the successful development of the nematode, all nutrients for embryogenesis and further growth need to be deposited into the egg. Nutrients are deposited in the egg as proteins and other macromolecules; the primarily stored proteins are yolk proteins, also known as vitellin (Almenara *et al.*, 2013). Vitellin is formed of a precursor protein vitellogenin, which is a large glyco-lipoprotein complex (Winter *et al.*, 1996), a distant relative of the apoB protein in humans (Baker, 1988a). In *Caenorhabditis elegans*, synthesis of vitellogenin occurs in the hermaphrodite intestine, which is then secreted into the body cavity and taken up by the oocytes by receptor-mediated endocytosis (Perez & Lehner, 2019); the protein is then deposited in yolk granules of the egg (Winter *et al.*, 1996). Deposits of yolk protein remain within the egg after embryogenesis and prior to hatching (Bossinger & Schierenberg, 2003).

Vitellogenin is believed to have the primary function of the transport of lipids and micronutrients (Hayward *et al.*, 2010) and amino acid provision to developing progeny (Winter *et al.*, 1996). Additionally, vitellogenin has also been shown to influence post-embryonic phenotypes, whereby limited supply of the protein can result in smaller sized offspring and sterility in *C. elegans* (Perez & Lehner, 2019). Some studies report that

1 vitellogenin also serves a protective role against oxidative stress (Ishii *et al.*, 2002) and
2 environmental protection from bacterial infections (Fischer *et al.*, 2013).
3

4
5 In *C. elegans*, vitellogenin proteins consist of four polypeptides: two large, 170 kDa
6 yp170A and 170 kDa yp170B, and two small, 115 kDa yp115 and 88 kDa yp88 (Perez &
7 Lehner, 2019). Six genes encoding vitellogenin have been characterised in *C. elegans* (*vit* 1-
8 6) (Spieth *et al.*, 1985). The peptide encoded by *vit*-6 is proteolytically cleaved prior to
9 entering the oocyte, to form yp115 and yp88 (Spieth *et al.*, 1991).
10
11

12
13
14
15
16
17 Vitellogenin proteins and the associated encoding genes have also been characterised
18 in some species of both vertebrates (fish and birds) and invertebrates (insects).
19
20
21 *Caenorhabditis elegans* has been the primary model for investigating vitellogenin in
22 nematodes; however, orthologous genes have been found in other free-living nematodes:
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

1 Vitellogenin is an important protein in yolk provision for the developing nematode
2 and possibly has a protective role; however, they have not been identified or characterised in
3
4 many species of cyst and root-knot nematodes. With the increasing availability of genomic
5
6 data for many of these economically important species, it is now possible to determine
7
8 presence of vitellogenin genes within their genomes and infer their phylogenetic
9
10 relationships.
11
12
13
14
15
16

17 **Materials and methods**

21 DATA ACQUISITION

22
23
24
25
26 Vitellogenin gene sequences from *C. elegans* (*vit-1*, *vit-2*, *vit-3*, *vit-4*, *vit-5*, and *vit-6*)
27
28 were obtained using wormbase.org (Table S1). These sequences were then used as the query
29
30 sequences for BLAST searches against the genomes of selected species of cyst nematodes
31
32 (*Globodera* and *Heterodera* spp.), root-knot nematodes (*Meloidogyne* spp.) and *C. briggsae*
33
34 on the parasite.wormbase.org server. BLAST searches were completed using tblastn, with a
35
36 maximum of 100 target sequences, expect threshold of 0.01, and a word size of 2. The
37
38 scoring parameters used were BLOSUM62 matrix and low complexity regions were filtered.
39
40 The result gave homologous regions and a corresponding predicted ‘overlapping gene’ in
41
42 each genome, these gene sequences were then selected from the database and collated for
43
44 sequence alignment.
45
46
47
48
49
50
51
52

53 GENE PREDICTION

1 The genome of *G. ellingtonae* was not available from the server of
2 parasite.wormbase.org; instead NCBI BLAST was used to determine homologous regions
3 within the species genome. Regions of the genome containing homologous sequences from
4 the resulting BLAST searches were selected and uploaded into the Galaxy web platform
5 (Afgan *et al.*, 2018); here the sequences were then inputted into the AUGUSTUS tool (Stanke
6 & Morgenstern, 2005; Stanke *et al.*, 2008). *Caenorhabditis elegans* was chosen as the
7 training set model organism, the protein sequence and coding sequence were then predicted
8 using both strands. The predicted vitellogenin genes were then selected by comparing the
9 sequence to homologous regions identified in the BLAST search.
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25

26 PHYLOGENETIC TREE CONSTRUCTION

27
28
29
30
31

32 All genes, predicted or database derived through BLAST searches, were then
33 uploaded to Pfam (Bateman *et al.*, 2004) to determine theoretical protein domains for each
34 protein sequence. Sequences which contained the same protein domains as *C. elegans*
35 vitellogenin genes were included in the sequence alignment. Sequences identified to have
36 homologous regions to the original queries but did not exhibit all domains shown in *C.*
37 *elegans* vitellogenin were excluded from these data. Sequences were uploaded into Molecular
38 Evolutionary Genetics Analysis version ten software (MEGA X) (Kumar *et al.*, 2018) and
39 pairwise alignment was performed using the MUSCLE (multiple sequence comparison by
40 log-expectation) function in the software (Edgar, 2004). The cluster method used was
41 UPGMA.
42
43
44
45
46
47
48
49
50
51
52
53
54
55

56 Phylogenetic trees were constructed using the maximum likelihood statistical method
57 in the MEGA X software, using the Jones-Taylor-Thornton substitution model, all sites
58
59
60
61
62
63
64
65

1 including gaps were used. For each tree constructed the test of phylogeny used was the
2 bootstrap method, this was replicated 100 times. Pairwise alignment scores were generated
3
4 using Jalview (Waterhouse *et al.*, 2009).
5
6
7
8

9 **Results**

10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
The result from the BLAST search indicated that there were homologous regions and
genes to *C. elegans* vitellogenin within PPN genomes that have not previously been
identified. Two genes from *G. ellingtonae* were predicted and were shown to be closely
related to two *G. rostochiensis* vitellogenin genes (Fig. 1). Predicted vitellogenin gene *G.*
ellingtonae 016 was shown to be most closely related to *G. rostochiensis* 2537, with a
pairwise alignment percentage identity of 98.7%, but only showed 91.96% identity with *G.*
pallida 0128. These three sequences differ in amino acid (aa) sequence length, 1963aa for *G.*
ellingtonae, 2322aa for *G. rostochiensis* 2537 and 2406aa for *G. pallida* 0128. The predicted
vitellogenin gene *G. ellingtonae* 060 was shown to be most closely related to *G. rostochiensis*
06238 with a percentage identity of 95.25% by pairwise alignment.

41 **PROTEIN DOMAINS IN VITELLOGENIN**

42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
Predicted vitellogenin protein domains illustrated in Figure 2, determined by Pfam,
show the approximate location of each protein domain within the genes. Three different
protein domains are exhibited in each gene: Vitellogenin_N (a lipoprotein amino terminal
region), vitellogenin open beta-sheet and von Willebrand factor type D domain. These
domains consistently appear in the same order for all genes but have a varied approximate
location. Figure 2 clearly illustrates that *C. elegans* and *C. briggsae* vitellogenin domains are

1 present at the similar estimated positions. Similarly, all *Heterodera* vitellogenin genes and *G.*
2 *ellingtonae* 016 have protein domains in comparable positions to each other. Another set of
3 genes with domains in similar positions include all *G. rostochiensis* vitellogenin genes and *G.*
4 *pallida* 9459. Two *M. incognita* vitellogenin genes (28653 and 30865) have almost identical
5 domain positions. Finally, many of the *M. javanica* predicted vitellogenin genes are
6 comparable to each other excluding *M. javanica* 29487 and 43787.
7
8
9
10
11
12
13
14
15
16
17

18 PHYLOGENETIC RELATIONSHIPS OF VITELLOGENIN 19 20 21 22 23 24

25 A phylogenetic tree was constructed to infer the relationships between vitellogenin
26 genes in cyst nematodes, root-knot nematodes and Rhabditida species, *C. elegans* and *C.*
27 *briggsae* (Fig. 1). The tree was split into two major groups, the PPN species and the
28 *Caenorhabditis* species; within the PPN group this was further divided into the root-knot and
29 cyst nematode species with a small outgroup consisting of *G. pallida* 0564, *H. glycines* 13367
30 and *H. schachtii* 19700. The cyst nematodes species gave mixed locations for each gene and
31 were not clustered based on their species. As shown in Figure 1, there were three smaller
32 clusters of PPN each containing two, three and five more closely related genes. Two
33 predicted genes from *H. schachtii* genome were included in the dataset, *H. schachtii* 007 was
34 shown to be most closely related to *H. glycines* 23424, whereas *H. schachtii* 19700 was in the
35 outgroup of the cyst nematode cluster.
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50

51 The root-knot nematode cluster within Figure 1, comprising *M. hapla*, *M. incognita*
52 and *M. javanica*, was split into two main groups, with a small outgroup including the only *M.*
53 *hapla* hypothesised vitellogenin gene to possess all three protein domains. Like the cyst
54
55
56
57
58
59
60
61
62
63
64
65

1 nematodes, different genes from each species were in mixed locations and did not form
2 clusters of individual species.
3

4 The divergence in genes between *C. elegans* and a closely related species, *C.*
5 *briggsae*, was considered. Figure 1 gives *C. elegans vit-6* and *C. briggsae* 006 as an outgroup
6 of the *Caenorhabditis* cluster, where the remaining genes are split into two groups.
7 Vitellogenin genes from *C. elegans*, *vit-1*, *vit-2* and *vit-3*, were more closely related to *C.*
8 *briggsae* 002 and 16767, whereas *C. elegans vit-4* and *vit-5* were more closely related to *C.*
9 *briggsae* 14203 and 14234.
10
11
12
13
14
15
16
17
18
19
20
21
22

23 Discussion

24
25
26
27
28 This paper highlights the presence of vitellogenin genes within the genomes of cyst
29 and root-knot nematodes that have not been previously extensively studied. The presence of
30 vitellogenin genes in the genomes of PPN signifies a requirement for yolk provision to the
31 eggs of the developing nematode. A significant proportion of the vitellogenin amino acid
32 sequence is conserved between the cyst nematode, root-knot nematode and *Caenorhabditis*
33 vitellogenin genes; however, areas of divergence remain. Conservation in parts of the amino
34 acid sequence may be due to the functional domains of the proteins encoded, whereas the
35 divergence of the genes may be a result of speciation events from a common ancestor.
36
37
38
39
40
41
42
43
44
45
46
47

48 Multiple paralogous vitellogenin genes are present within many species of both
49 vertebrates and invertebrates, of differing amino acid composition (Smolenaars *et al.*, 2007).
50 In *C. elegans* these genes encode for different subunits of a larger protein complex, which
51 may allow for varying function of the encoded proteins (Perez & Lehner, 2019); it is possible
52 that multiple genes in PPN are required to form similar complexes. Data presented here
53
54
55
56
57
58
59
60
61
62
63
64
65

1 illustrate the relationship between these paralogous genes, with each species having differing
2 numbers of genes. Only one vitellogenin gene from the genome of *M. hapla* was identified
3 that contained all three protein domains, whereas the genome of *M. incognita* has seven
4 genes. This may be due to the relative sizes of their genomes, *M. hapla* genome contains 54
5 million base pairs (Opperman *et al.*, 2008), whilst the *M. incognita* genome contains 86
6 million base pairs (Abad *et al.*, 2008).
7
8
9
10
11
12
13

14 The genome of *G. ellingtonae* contains two not previously predicted vitellogenin
15 genes, which are both most closely related to two genes from *G. rostochiensis*. *Globodera*
16 *ellingtonae* has been shown to have similar behaviour and requirements for hatching and
17 development to *G. rostochiensis*, but shows some molecular diversity (Hesse *et al.*, 2021).
18 Like *G. rostochiensis*, *G. ellingtonae* is a pathogen of certain cultivars of potato but its
19 pathogenicity has been inconsistent (Zasada *et al.*, 2019). The inclusion of this recently
20 discovered potato cyst nematode, *G. ellingtonae*, sought to further evaluate the nematode
21 species against other members of the *Globodera* family.
22
23
24
25
26
27
28
29
30
31
32
33

34 As vitellogenin plays a key role in the development of the nematode, it could be
35 anticipated that it would have a relatively conserved sequence, especially between closely
36 related species. This has also been illustrated in the comparison of *C. elegans* with *C.*
37 *briggsae* where sequences were shown to be 85% identical in coding regions (Zucker-
38 Aprison & Blumenthal, 1989); therefore, vitellogenin genes in *C. elegans* and *C. briggsae* are
39 highly conserved. The genome of *C. elegans* contains six vitellogenin genes whereas *C.*
40 *briggsae* only contains five (Zucker-Aprison & Blumenthal, 1989). Perez & Lehner (2019)
41 suggest that the *vit-4* gene is missing from *C. briggsae* due to a duplication between *vit-3* and
42 *vit-4*. Contrary to this, data presented here indicate that there are two genes closely related to
43 *C. elegans vit-4* and *vit-5* in *C. briggsae*, but only two genes closely related to *C. elegans vit-*
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1, *vit-2* and *vit-3*. This result suggests that either *vit-1*, *vit-2* or *vit-3* may be missing from *C. briggsae*, not *vit-4* as originally predicted.

Vitellogenin genes have been shown to be upregulated in *G. pallida*. In supplementary data provided by Cotton *et al.* (2014), three *G. pallida* vitellogenin genes were shown to be upregulated in females (at 21, 28 and 35 dpi) compared to the early parasitic life-stage (7 and 14 dpi). The three genes shown to be upregulated are presented here and form part of the main cyst nematode cluster in the phylogenetic analyses. The gene *G. pallida* 0564, which forms part of an outgroup from the PPN cluster, is not shown to be expressed in that study; it is possible that this gene was not identified, or this gene is not expressed. Currently, no vitellogenin genes have been characterised in *G. rostochiensis*.

The genomes of species within the cyst nematode genus *Heterodera* contain vitellogenin genes; this study highlights their presence within the genomes of *H. schachtii* and *H. glycines*. Two *H. schachtii* genes are shown to be closely related to those in *H. glycines*; however, a third *H. glycines* gene appears to be more divergent. This result supports the findings of Singh *et al.* (2020), who generated a phylogenetic consensus tree containing 33 different populations of *Heterodera* spp., finding that *H. schachtii* and *H. glycines* were closely related. In a study that sequences the *H. schachtii* genome, data provided in the supplementary information listed two proteins that were annotated as containing a lipoprotein amino terminal region, von Willebrand factor type D domain and a domain of unknown function, whilst also indicating that they were highly expressed at 24 dpi (Siddique *et al.*, 2021). These two proteins have been incorporated into Figure 1 as HS_19700 and HS_007; establishing that they also have a vitellogenin open beta-sheet domain, which was not previously annotated.

Masler (1999) partially characterised two vitellogenin proteins in *H. glycines* with similar molecular weights (180 and 190 kDa) to the two larger protein subunits of *C. elegans*

1 (170 kDa yp170A and 170 kDa yp170B). Figure 1 presents three *H. glycines* hypothesised
2 vitellogenin genes and it is possible that these genes encode for the ‘major egg proteins’ as
3 described (Masler, 1999). The *H. glycines* transcriptome of effectors gave only one predicted
4 vitellogenin gene sequence, provided in the supplementary information (Gardner *et al.*,
5 2018). Again, it may be possible that not all vitellogenin genes present in the genome are
6 expressed to produce the vitellogenin proteins.
7

8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Root-knot nematodes are the most damaging nematode crop pests, so it was important to include some species of this genus. The relationship between 13 vitellogenin genes from root-knot nematode species has been illustrated, including one gene from the genome of *M. hapla*, five from *M. javanica* and seven from the genome of *M. incognita*. Currently, no studies have highlighted the presence of vitellogenin within the genomes of *M. hapla* or *M. javanica*. In previous analysis of the *M. incognita* genome, no vitellogenin encoding genes were identified (Abad *et al.*, 2008); however, in supplementary data provided by Bellafiore *et al.* (2008), five vitellogenin genes were found in the secreted effectors of female pharyngeal glands using mass spectrometry. A similar study of the effectors produced by female pharyngeal glands in *G. rostochiensis* resulted in no vitellogenin genes being detected (Maier *et al.*, 2012). The result from Bellafiore *et al.* (2008) is unexpected, as it is understood from *C. elegans* that vitellogenin genes are primarily expressed in the hermaphrodite intestine and then transported into the oocyte by receptor-mediated endocytosis (Perez & Lehner, 2019). However, it has now been shown that the *C. elegans* hermaphrodite vents vitellogenin as part of the yolk after egg laying ceases, for further nutritional provision for the developing nematode (Kern *et al.*, 2021). It may be possible that root-knot nematodes exhibit a similar behaviour when producing an egg mass, which primarily consists of glycoproteins (Sharon & Spiegel, 1993).

1 Protein domains present in *C. elegans* vitellogenin appear within the identified
2 vitellogenin genes of cyst and root-knot nematodes. Using Pfam to predict the location of
3 protein domains, *C. elegans*, *C. briggsae*, cyst and root-knot nematode vitellogenin genes
4 were shown to contain three protein domains: Vitellogenin_N, vitellogenin open beta-sheet
5 and von Willebrand factor type D domain. In BLAST searches of PPN genomes, some genes
6 contained only one or two of these domains, with a sequence length that was much shorter
7 than other vitellogenin genes within the species; therefore, these results were not included in
8 the analysis. Protein domain Vitellogenin_N is a lipoprotein amino terminal region that is
9 predicted to be involved in lipid transport and is a domain present in the vitellogenin of
10 certain insect, crustacean and nematode species (Smolenaars *et al.*, 2007). Nematode
11 vitellogenin have been shown to contain amphipathic β -strands, which may also function in
12 the transport of additional lipids (Smolenaars *et al.*, 2007). These strands may then form the
13 vitellogenin open beta-sheet domain. Towards the C-terminus of the vitellogenin gene in
14 nematodes, the protein domain von Willebrand type D is present. This domain has shown
15 adhesive properties and has been hypothesised to be the binding site for membrane receptors
16 on the oocyte (Baker, 1988b).

17 Whilst the presence of vitellogenin genes within certain PPN species has been
18 determined, a complete understanding of their expression and function is yet to be fully
19 elucidated. Each nematode species contains a different number of vitellogenin genes, even
20 between closely related species. There is a possibility that the variance in the number of
21 genes may be due to differences in expression, with some genes not being expressed. Further
22 analysis of the stage- and tissue-specific expression of vitellogenin would help to understand
23 further their presence and role in PPN. Vitellogenin genes in selected cyst and root-knot
24 nematode species contain the same protein domains as those in *C. elegans* so could be
25 hypothesised to exhibit a similar function; however, this requires further investigation.

1 Masler (1999) has partially characterised vitellogenin proteins in *H. glycines*; this work could
2 be expanded to characterise the proteins in other PPN species. Vitellogenin plays a key role
3 in the transport of lipids and nutritional provision to the developing nematode; if this function
4 could be interrupted this may provide a novel route for control of PPN. Additionally,
5 vitellogenin has been shown to have a role in protection against environmental stress in *C.*
6
7
8
9
10
11
12 *elegans* (Fischer *et al.*, 2013). This has been demonstrated by the knockdown of *vit-6* gene by
13
14 RNA-interference, which increased the nematodes susceptibility to *Photorhabdus*
15
16 *luminescens*, reducing its lifespan (Fischer *et al.*, 2013). If vitellogenin inhibited in PPN, this
17
18 may provide a novel method of control by potentially increasing the nematode's
19
20
21
22
23 susceptibility to environmental stress.

24 This study set out to highlight the presence of and predict vitellogenin genes within
25
26 the genomes of some economically important PPN. The genes encoding the vitellogenin
27
28 protein in *C. elegans* were found to be homologous to sequences in the genomes of cyst and
29
30 root-knot nematodes and could be used to infer relationships between the PPN species.
31
32
33 Additionally, the protein domains of vitellogenin genes that appear in *C. elegans* were also
34
35 present in the PPN. This study contributes to understanding of vitellogenin in PPN; however,
36
37
38 further investigation needs to be completed to fully characterise the proteins and understand
39
40
41 their biological role. By determining the presence of vitellogenin genes within the genomes
42
43 of some economically important PPN, the relationship between these genes could be inferred
44
45
46 and this information could help form a basis for further investigation of vitellogenin in PPN.
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

References

- 1
2
3
4
5 ABAD, P., GOUZY, J., AURY, J.M., CASTAGNONE-SERENO, P., DANCHIN, E.G., DELEURY, E.,
6
7 PERFUS-BARBEOCH, L., ANTHOUARD, V., ARTIGUENAVE, F., BLOK, V.C. *et al.* (2008).
8
9 Genome sequence of the metazoan plant-parasitic nematode *Meloidogyne*
10
11 *incognita*. *Nature Biotechnology* 26, pp. 909-915. DOI: 10.1038/nbt.1482
12
13
14 AFGAN, E., BAKER, D., BATUT, B., VAN DEN BEEK, M., BOUVIER, D., ČECH, M., CHILTON, J.,
15
16 CLEMENTS, D., CORAOR, N., GRÜNING, B.A. *et al.* (2018). The Galaxy platform for
17
18 accessible, reproducible, and collaborative biomedical analyses: 2018 update. *Nucleic*
19
20 *Acids Research* 46, pp. 537-544. DOI:10.1093/nar/gky379
21
22
23
24 ALMENARA, D.P., DE MOURA, J.P., SCARABOTTO, C.P., ZINGALI, R.B. & WINTER, C.E. (2013).
25
26 The molecular and structural characterization of two vitellogenins from the free-living
27
28 nematode *Oscheius tipulae*. *PloS One* 8, e53460. DOI: 10.1371/journal.pone.0053460
29
30
31 BAKER, M.E., (1988a). Is vitellogenin an ancestor of apolipoprotein B-100 of human low-
32
33 density lipoprotein and human lipoprotein lipase?. *Biochemical Journal* 255, pp.
34
35 1057-1060. DOI: 10.1042/bj2551057.
36
37
38
39 BAKER, M.E., (1988b). Invertebrate vitellogenin is homologous to human von Willebrand
40
41 factor. *Biochemical Journal* 256, p. 1059. DOI: 10.1042/bj2561059
42
43
44 BATEMAN, A., COIN, L., DURBIN, R., FINN, R.D., HOLLICH, V., GRIFFITHS-JONES, S., KHANNA,
45
46 A., MARSHALL, M., MOXON, S., SONNHAMMER, E.L. *et al.* (2004). The Pfam protein
47
48 families database. *Nucleic Acids Research* 32, pp. 138-141. DOI: 10.1093/nar/gkh121
49
50
51 BELLAFFIORE, S., SHEN, Z., ROSSO, M.N., ABAD, P., SHIH, P. AND BRIGGS, S.P. (2008). Direct
52
53 identification of the *Meloidogyne incognita* secretome reveals proteins with host cell
54
55 reprogramming potential. *PLoS Pathogens* 4, e1000192. DOI:
56
57 10.1371/journal.ppat.1000192
58
59
60
61
62
63
64
65

- 1 BOSSINGER, O. & SCHIERENBERG, E. (2003). The use of fluorescent marker dyes for studying
2 intercellular communication in nematode embryos. *International Journal of*
3 *Developmental Biology* 40, pp. 431-439.
4
5
6
7 COTTON, J.A., LILLEY, C.J., JONES, L.M., KIKUCHI, T., REID, A.J., THORPE, P., TSAI, I.J.,
8
9
10 BEASLEY, H., BLOK, V., COCK, P.J. *ET AL.* (2014). The genome and life-stage specific
11 transcriptomes of *Globodera pallida* elucidate key aspects of plant parasitism by a
12 cyst nematode. *Genome Biology* 15, pp. 1-17. DOI: 10.1186/gb-2014-15-3-r43
13
14
15
16
17 EDGAR, R.C. (2004). MUSCLE: a multiple sequence alignment method with reduced time
18 and space complexity. *BMC Bioinformatics* 5, pp. 1-19. DOI: 10.1186/1471-2105-5-
19
20
21 113
22
23
24 FISCHER, M., REGITZ, C., KULL, R., BOLL, M. & WENZEL, U. (2013). Vitellogenins increase
25 stress resistance of *Caenorhabditis elegans* after *Photorhabdus luminescens* infection
26 depending on the steroid-signaling pathway. *Microbes and Infection* 15, pp. 569-578.
27
28
29
30
31
32
33
34 GARDNER, M., DHROSO, A., JOHNSON, N., DAVIS, E.L., BAUM, T.J., KORKIN, D. & MITCHUM,
35
36
37 M.G. (2018). Novel global effector mining from the transcriptome of early life stages
38 of the soybean cyst nematode *Heterodera glycines*. *Scientific Reports* 8, pp. 1-15.
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
- HAYWARD, A., TAKAHASHI, T., BENDENA, W.G., TOBE, S.S. & HUI, J.H. (2010). Comparative
genomic and phylogenetic analysis of vitellogenin and other large lipid transfer
proteins in metazoans. *FEBS Letters* 584, pp. 1273-1278. DOI:
10.1016/j.febslet.2010.02.056
- HESSE, C.N., MORENO, I., PARDO, O.A., FUENTES, H.P., GRENIER, E., DANDURAND, L.M. &
ZASADA, I.A. (2021). Characterization of Populations from Chile Utilizing Whole

1 Genome Sequencing. *Journal of Nematology* 53, pp. 1-9. DOI: 10.21307/jofnem-
2 2021-088
3

4 ISHII, N., GOTO, S. & HARTMAN, P.S. (2002). Protein oxidation during aging of the nematode
5 *Caenorhabditis elegans*. *Free Radical Biology and Medicine* 33, pp. 1021-1025. DOI:
6
7 10.1016/S0891-5849(02)00857-2
8
9

10 JONES, J.T., KUMAR, A., PYLYPENKO, L.A., THIRUGNANASAMBANDAM, A., CASTELLI, L.,
11
12 CHAPMAN, S., COCK, P.J., GRENIER, E., LILLEY, C.J., PHILLIPS, M.S. *et al.* (2009).
13
14 Identification and functional characterization of effectors in expressed sequence tags
15
16 from various life cycle stages of the potato cyst nematode *Globodera*
17
18 *pallida*. *Molecular Plant Pathology* 10, pp. 815-828. DOI: 10.1111/j.1364-
19
20 3703.2009.00585.x
21
22
23
24

25 JONES, J.T., HAEGEMAN, A., DANCHIN, E.G., GAUR, H.S., HELDER, J., JONES, M.G., KIKUCHI,
26
27 T., MANZANILLA-LÓPEZ, R., PALOMARES-RIUS, J.E., WESEMAEL, W.M. *et al.* (2013).
28
29 Top 10 plant-parasitic nematodes in molecular plant pathology. *Molecular Plant*
30
31 *Pathology* 14, pp. 946-961. DOI: 10.1111/mpp.12057
32
33
34

35 KERN, C.C., TOWNSEND, S., SALZMANN, A., RENDELL, N.B., TAYLOR, G.W., COMISEL, R.M.,
36
37 FOUKAS, L.C., BÄHLER, J. & GEMS, D. (2021). *C. elegans* feed yolk to their young in a
38
39 form of primitive lactation. *Nature Communications* 12, pp. 1-11. DOI:
40
41 10.1038/s41467-021-25821-y
42
43
44

45 KUMAR, S., STECHER, G., LI, M., KNYAZ, C., & TAMURA, K. (2018). MEGAX: molecular
46
47 evolutionary genetics analysis across computing platforms. *Molecular Biology and*
48
49 *Evolution* 35, p. 1547. DOI: 10.1093/molbev/msy096
50
51

52 KUMAR, Y. & YADAV, B.C. (2020). Plant-parasitic nematodes: Nature's most successful plant
53
54 parasite. *International Journal of Research and Review* 7, pp. 379-386.
55
56
57
58
59
60
61
62
63
64
65

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
- MAIER, T.R., HEWEZI, T., PENG, J. & BAUM, T.J. (2013). Isolation of whole esophageal gland cells from plant-parasitic nematodes for transcriptome analyses and effector identification. *Molecular Plant-Microbe Interactions* 26, pp. 31-35. DOI: 10.1094/MPMI-05-12-0121-FI
- MASLER, E.P. (1999). Detection and partial characterization of egg polypeptides from *Heterodera glycines*. *Journal of Nematology* 31, p. 305-311.
- MKANDAWIRE, T.T., GRENCIS, R.K., BERRIMAN, M. & DUQUE-CORREA, M.A. (2021). Hatching of parasitic nematode eggs: a crucial step determining infection. *Trends in Parasitology* 38, pp. 174-187. DOI: 10.1016/j.pt.2021.08.008
- NICOL, J.M., TURNER, S.J., COYNE, D.L., NIJS, L.D., HOCKLAND, S. & MAAFI, Z.T. (2011). Current nematode threats to world agriculture. In: Jones, J., Gheyson, G. & Fenoll, C. (Eds). *Genomics and Molecular Genetics of Plant-Nematode Interactions*. Dordrecht, Netherlands, Springer, pp. 21-43. DOI: 10.1007/978-94-007-0434-3_2
- OPPERMAN, C.H., BIRD, D.M., WILLIAMSON, V.M., ROKHSAR, D.S., BURKE, M., COHN, J., CROMER, J., DIENER, S., GAJAN, J., GRAHAM, S. *et al.* (2008). Sequence and genetic map of *Meloidogyne hapla*: A compact nematode genome for plant parasitism. *Proceedings of the National Academy of Sciences* 105, pp. 14802-14807. DOI: 10.1073/pnas.0805946105
- PEREZ, M.F. & LEHNER, B. (2019). Vitellogenins- yolk gene function and regulation in *Caenorhabditis elegans*. *Frontiers in Physiology* 10, p. 1067. DOI: 10.3389/fphys.2019.01067
- SHARON, E. & SPIEGEL, Y. (1993). Glycoprotein characterization of the gelatinous matrix in the root-knot nematode *Meloidogyne javanica*. *Journal of Nematology* 25, p.585-589.
- SIDDIQUE, S., RADAKOVIC, Z.S., HILTL, C., PELLEGRIN, C., BAUM, T.J., BEASLEY, H., CHITAMBO, O., CHOPRA, D., DANCHIN, E.G., GRENIER, E. *et al.* (2021). The genome

1 and life stage-specific transcriptomes of a plant-parasitic nematode and its host reveal
2 susceptibility genes involved in trans-kingdom synthesis of vitamin B. *bioRxiv*. DOI:
3
4 10.1101/2021.10.01.462558
5
6

7 SINGH, P.R., KARSSSEN, G., COUVREUR, M. & BERT, W. (2020). Morphological and molecular
8
9 characterization of n. sp. (Nematoda: Heteroderidae) from Gran Canaria, Canary
10
11 Islands. *Journal of Nematology* 52, pp. 1-14. DOI: 10.21307/jofnem-2020-098
12
13

14 SMOLENAARS, M.M., MADSEN, O., RODENBURG, K.W. & VAN DER HORST, D.J. (2007).
15
16 Molecular diversity and evolution of the large lipid transfer protein
17
18 superfamily. *Journal of Lipid Research* 48, pp. 489-502. DOI: 10.1194/jlr.R600028-
19
20 JLR200
21
22

23
24 SPIETH, J., DENISON, K., ZUCKER, E. & BLUMENTHAL, T., (1985). The nucleotide sequence of
25
26 a nematode vitellogenin gene. *Nucleic Acids Research* 13, pp. 7129-7138. DOI:
27
28 10.1093/nar/13.19.7129
29
30

31 SPIETH, J., NETTLETON, M., ZUCKER-APRISON, E., LEA, K. & BLUMENTHAL, T. (1991).
32
33 Vitellogenin motifs conserved in nematodes and vertebrates. *Journal of Molecular*
34
35 *Evolution* 32, pp. 429-438. DOI: 10.1007/BF02101283
36
37

38
39 STANKE, M., DIEKHANS, M., BAERTSCH, R. & HAUSSLER, D. (2008). Using native and
40
41 syntenically mapped cDNA alignments to improve de novo gene
42
43 finding. *Bioinformatics* 24, pp. 637-644. DOI: 10.1093/bioinformatics/btn013
44
45

46 STANKE, M. & MORGENSTERN, B. (2005). AUGUSTUS: a web server for gene prediction in
47
48 eukaryotes that allows user-defined constraints. *Nucleic Acids Research* 33, pp. 465-
49
50 467. DOI: 10.1093/nar/gki458
51
52

53 VIEIRA, P., EVES-VAN DEN AKKER, S., VERMA, R., WANTOCH, S., EISENBACK, J.D. & KAMO,
54
55 K. (2015). The *Pratylenchus penetrans* transcriptome as a source for the development
56
57 of alternative control strategies: mining for putative genes involved in parasitism and
58
59

1 evaluation of in planta RNAi. *PloS One* 10, e0144674. DOI:
2 10.1371/journal.pone.0144674
3

4 WATERHOUSE, A.M., PROCTER, J.B., MARTIN, D.M., CLAMP, M. & BARTON, G.J. (2009).
5 Jalview Version 2--a multiple sequence alignment editor and analysis
6
7 workbench. *Bioinformatics* 25, pp. 1189-1191. DOI: 10.1093/bioinformatics/btp033
8
9

10 WINTER, C.E., PENHA, C. & BLUMENTHAL, T. (1996). Comparison of a vitellogenin gene
11
12 between two distantly related rhabditid nematode species. *Molecular Biology and*
13
14 *Evolution* 13, pp. 674-684. DOI: 10.1093/oxfordjournals.molbev.a025628
15
16
17

18 ZASADA, I. A., INGHAM, R. E., BAKER, H., & PHILLIPS, W. S. (2019). Impact of *Globodera*
19
20 *ellingtonae* on yield of potato (*Solanum tuberosum*). *Journal of Nematology* 51. DOI:
21
22 10.21307/JOFNEM-2019-073
23
24
25

26 ZUCKER-APRISON, E. & BLUMENTHAL, T. (1989). Potential regulatory elements of nematode
27
28 vitellogenin genes revealed by interspecies sequence comparison. *Journal of*
29
30 *Molecular Evolution* 28, pp. 487-496. DOI: 10.1007/BF02602929
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

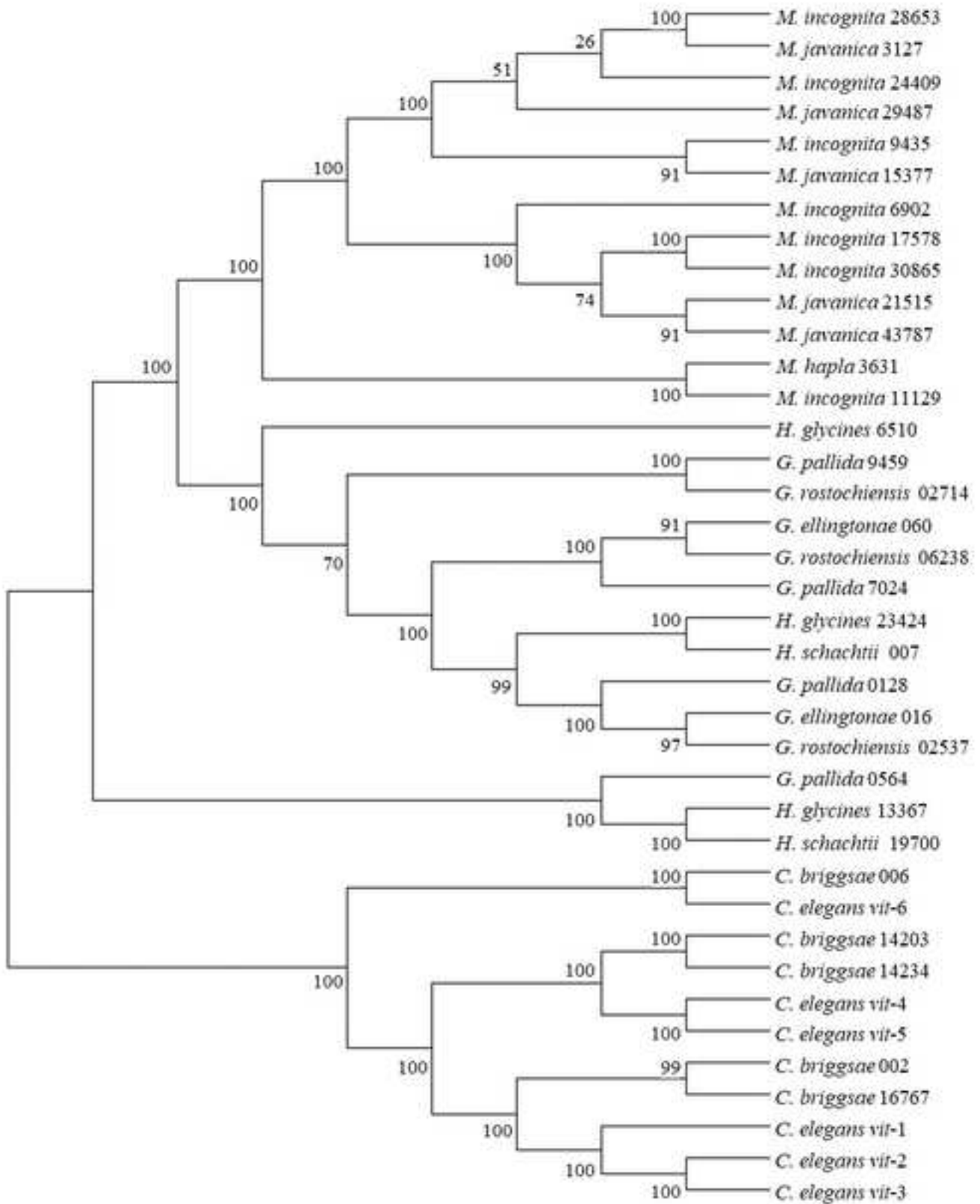


Figure 2. Protein Domain Schematic

[Click here to access/download;Figure;Figure2. Protein domans.jpg](#)



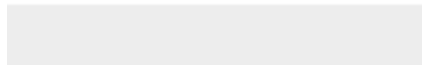


Click here to access/download
Supplementary material
Figure legend.rtf





Click here to access/download
Supplementary material
Supplementary Table S1.rtf





Click here to access/download
Supplementary material
Cover Letter.rtf

