Plant Molecular Biology 12: 597–600, 1989

Complete nucleotide sequence of the two homeologous tobacco nitrate reductase genes

H. Vaucheret, J. Kronenberger, P. Rouzé and M. Caboche*
*Laboratoire de Biologie Cellulaire, INRA, 78026 Versailles Cedex, France (*author for correspondence)*

We have previously described the molecular cloning and characterization of one of the two tobacco nitrate reductase structural genes, nia-2 [1]. In this report we compare the sequence of nia-2 with that of nia-1, which has also been cloned (Fig. 1). The size of exons and position of introns are identical in both genes. The four exons of 1012, 141, 233 and 1326 bp respectively encode a 904 amino acid polypeptide of 101.7 kDa for the *Nicotiana tomentosiformis*-derived nia-1 gene and of 101.8 kDa for the *Nicotiana sylvestris*-derived nia-2 gene. The homology between exons of the two genes is very high at the nucleotide level (97%) and at the amino acid level (98%). Introns display a lower degree of homology. They are always longer in the nia-2 gene than in the nia-1 gene (594 nt versus 275 nt for IVS1, 1298 nt versus 912 nt for IVS2, 788 nt versus 652 nt for IVS3) with extra sequences localized in the vicinity of their 5′ ends, and correlatively more conserved 3′ ends (about 70% of homology in the colinear part of the introns). An extensive homology of the 5′ non-coding region sequences of the two genes was detected up to 800 bp upstream from the transcription initiation sites and should include all the regulatory elements of the promoter. A highly repeated (TA)n motif was found in the 3′ non coding sequences. It lies 360 and 650 bp downstream the TAA stop codon in nia-1 and nia-2 respectively, TA being repetead 27 times in the structure of the nia-1 gene and 14 times in the structure of the nia-2 gene. Such a sequence has recently been related to the rate of degradation of the mRNA [2]. Significant homology can be found a long distance away from the transcribed region, 1150 bp and 850 bp downstream the TAA stop codon for the nia-1 and nia-2 genes respectively (data not shown).

Acknowledgements

Sequence data analyse were performed with the BISANCE package using the CITI2 computer facilities, granted by the French 'Ministere de la Recherche et de la Technologie'.

References

1. Vaucheret H, Vincentz M, Kronenberger J, Caboche M, Rouze P: Molecular cloning and characterization of the two homeologous genes coding for nitrate reductase in tobacco. Mol Gen Genet, in press (1989).
2. Wilson T, Treisman R: Removal of poly(A) and consequent degradation of c-fos mRNA facilitated by 3′ AU-rich sequences. Nature 336: 396–399 (1988).

Fig. 1. Nucleotides have been numbered from the transcription initiation sites on the two genes. Upper lane: nia-1, lower lane: nia-2. Identical sequences for the two genes are represented by dots in the upper lane and deletions by dashes. Non-coding sequences are printed in lower-case letters. Coding sequences, in capital letters, have been translated using the single letter aminoacid code. For divergent amino acid sequences the first symbol refers to the nia-1 gene. CAAT and TATA boxes, the transcription initiation and termination putative sites and translation start codon are indicated by asterisks.

The nucleotide sequence data reported will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession numbers X14058 Nia 1 and X14059 Nia 2.

```
-865 Nia-1 .ctt.gagtaatgtat.catttaagagc.at.ta..t..a........c.c...............agc....c......c......c.ta...
-965 Nia-2 tacatacaagggcgcgaataaactttttttaaagtaaatgtatatgaacttgcaatgaaagaggaccttaacttgtttgtctttgttgctttc

-771 ...........gt....a.................a...a..c........taacgatatcaaagaaaac..............................
-871 tgcaaatttcaccttaacagcccatttgagattgatttagttagttataac-------------------aattagttaaatgcttgtgtaatttgaaga

-671 .tt.t..........g............g----------------........................................a...........a.t
-790 aaatatttggacgtgctcgctgaaaacattatactcctatataatagaaatactttctgaaaagttggtcttgttcaaaaacgta-taagagagttggt-

-587 ...t.t.g....a.......----......ga..---........ga....c..g.........t...g.t....a.......t...............g
-692 cttct-cataaatagtcactagctttctg--attttttttcactttctatatcacgtaaataggtactcaaatttgatatttacaccaaacaaatgaaaa

-494 ..aa........g.c........a..............a..g.........t...t.....t..tt........a.......c........-.g......
-595 taggatatgtgt-t-tttcatacgtatatttatctatcgtacttaatgatacatacatatacata-taaccttactttttgattactaaaaatttaatta

-395 ...............g..............t...............------------------------------------------------------
-498 tatttaatttgggtaaatatcagatgccacaaaacatttacctagccactgtttttgactactaaaaatttaattatgtttagcttgggtaaatatcaga

-349 ---------------------------..........t............................................................a.
-398 tgtcactaaacattttacctagccattcctccgaaaagaaattgagaaggaaattagagttagtggagccataataatgtttaatgtgaccataactcgg

-276 .--------------------.............gtc.............................................tcc...............
-298 tgaaaaccacggcaagaataagaaacagctgttaaggctaaccaacagctgcatatctttaagccatttgctattaccccaacatcgcatcttcctctga

-196 .................................a................----.g..................t....******...............
-198 tcccgaccctacgggcgtaaaaagtgtaaatcgttagaattgttttatttattttatgatgtcactatttttt-aaaatcaaaattaaattggggtgtcg

-100 ..............ca................gg.............t............t.......*******.........................
- 99 atttttttgggtcctgcttatgtatagtatggc-gctatggaggcactgagagagtccgaaacgtttctatataaggccaccccacgcattcacaaactt

   1 *...--------------.............ta...................................................----------------
   1 cgttcccaaacagaacaagaaaatcaaatctcggagagagagagagagaaatatttgagagagaaatacagaaaatctctcttccttctttcctttttt

  70 ------------------..-................. *** ... ... ... ... ... ... ... ... ... ... ... A.. ... ...
 101 tttcaatccccattcatattctttttttagaataatct ATG GCG GCA TCT GTC GAA AAC AGG CAG TTC AGT CAC CTA GAA GCC
   1                                         M   A   A   S   V   E   N   R   Q   F   S   H  I/L  E   A

 135 ... ... ... ... ... ... ... ..T ... ... ... ... ... ... ... ... ... ... ... ... T.. ... ... ... ...
 184 GGT TTA TCC CGG TCT TTC AAG CCC CGG TCT GAT TCC CCG GTT CGT GGC TGC AAC TTC CCT CCG CCC AAC AGT ACT
  16  G   L   S   R   S   F   K   P   R   S   D   S   P   V   R   G   C   N   F   P  S/P  P   N   S   T

 210 ... ... ... ... ... ... ... ... ... ... .T. ... ... ... ... ... ... ... ... ... ... ... ... ... ...
 259 AAT TTC CAA AAG AAA CCA AAT TCC ACC ATT TAC CTT GAT TAC TCG TCG AGT GAA GAC GAC GAT GAT GAT GAC GAA
  41  N   F   Q   K   K   P   N   S   T   I  F/Y  L   D   Y   S   S   S   E   D   D   D   D   D   D   E

 285 ... ... ... ... ... ... ... ..C ... ... ... ... ... ..A ... ... ... ... ... ... ... .GC ... ... ...
 334 AAA AAT GAG TAC CTT CAA ATG ATT AAA AAA GGG AAT TCA GAG TTA GAG CCA TCT GTT CAT GAC ACT AGG GAC GAA
  66  K   N   E   Y   L   Q   M   I   K   K   G   N   S   E   L   E   P   S   V   H   D  S/T  R   D   E

 360 ... ... ... ... ... ... ... ... ... ... ... ... T.. ... ... ... ... ... ... ... ... ... ... ... ...
 409 GGT ACC GCT GAT AAT TGG ATT GAA CGC AAC TTT TCC ATG ATT CGT CTC ACC GGA AAG CAT CCA TTT AAC TCC GAA
  91  G   T   A   D   N   W   I   E   R   N   F   S  L/M  I   R   L   T   G   K   H   P   F   N   S   E

 435 ..G ... ... ... ..T ... ... ... ... ..T ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
 484 CCA CCG TTG AAC CGG CTC ATG CAC CAC GGC TTT ATC ACA CCG GTC CCA CTT CAT TAC GTT CGT AAC CAT GGA CCG
 116  P   P   L   N   R   L   M   H   H   G   F   I   T   P   V   P   L   H   Y   V   R   N   H   G   P

 510 ... ... ... ... ..A ... ... ... ... ... ... ... ... ... ... ... ... ..A ... ... ... ... ... ... ...
 559 GTT CCC AAG GGC ACG TGG GAT GAC TGG ACC GTG GAA GTC ACG GGA CTA GTG AAG CGT CCT ATG AAA TTC ACA ATG
 141  V   P   K   G   T   W   D   D   W   T   V   E   V   T   G   L   V   K   R   P   M   K   F   T   M

 585 ... ... ... ... ... ... ... ... .CC ... ... ... ... ... ... ... ... ... ... ... ..C ... ... ... ...
 634 GAC CAG TTG GTT AAC GAA TTC CCT TGT AGA GAA TTG CCC GTT ACG CTT GTT TGT GCT GGC AAT CGA AGG AAA GAA
 166  D   Q   L   V   N   E   F   P  S/C  R   E   L   P   V   T   L   V   C   A   G   N   R   R   K   E

 660 ... ... ... ... ... ... ... ... ... ... ... ... ..T ... ... ... ... ... ... ..T G.. ... ... ... ...
 709 CAG AAC ATG GTT AAA CAA ACC ATT GGT TTC AAC TGG GGC GCC GCT GCC GTT TCA ACA ACG ATA TGG CGC GGG GTA
 191  Q   N   M   V   K   Q   T   I   G   F   N   W   G   A   A   A   V   S   T   T  V/I  W   R   G   V

 735 ... ..A ... ... ... T.. ... ... .A. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
 784 CCC CTC CGC GCT TTG CTA AAA CGG TGC GGT GTT TTT AGC AAG AAT AAA GGG GCG CTT AAT GTT TGC TTC GAA GGA
 216  P   L   R   A   L   L   K   R  Y/C  G   V   F   S   K   N   K   G   A   L   N   V   C   F   E   G

 810 ... ... ... ... ... ... ..C ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
 859 GCT GAT GTG TTG CCC GGA GGT GGT GGT TCA AAG TAT GGA ACC AGC ATT AAG AAG GAA TTT GCA ATG GAT CCA GCA
 241  A   D   V   L   P   G   G   G   G   S   K   Y   G   T   S   I   K   K   E   F   A   M   D   P   A

 885 ... ... ... ..A A.. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
 934 CGA GAT ATC ATC GTA GCC TAC ATG CAG AAC GGA GAA AAA TTG GCA CCC GAC CAC GGG TTT CCA GTA CGA ATG ATA
 266  R   D   I   I  I/V  A   Y   M   Q   N   G   E   K   L   A   P   D   H   G   F   P   V   R   M   I

 960 ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
1009 ATT CCA GGA TTC ATT GGA GGA AGA ATG GTG AAA TGG ATA AAG AGG ATT ATA GTC ACC ACC CAA GAA TCA GAC AGC
 291  I   P   G   F   I   G   G   R   M   V   K   W   I   K   R   I   I   V   T   T   Q   E   S   D   S

1035 ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ..T ... . ...-------
1084 TAT TAT CAT TTC AAG GAC AAT AGA GTT CTT CCT CCC CAT GTT GAT GCT GAA CTT GCA AAT ACC GAA G gtacgtaccg
 316  Y   Y   H   F   K   D   N   R   V   L   P   P   H   V   D   A   E   L   A   N   T   E

//// ----------------------------------------------------------------------------------------------------
1161 taactatttcaatttattactccatttgttccaatttatgtgaacctatttcctttttggtccgttcaaaaagaatgaaccctttctaaatttggtaac

//// ----------------------------------------------------------------------------------------------------
1261 aatttagcttaaacttacaacttcacccttaatgagaaactttataaccacacaaataccctggggcccatttggacttgtttaggtcgacaaattcca
```

```
1105 ------------------------------------------------------------------------............t...c-..-.......
1361 aaagtttatttttttcttaaacttcgtgctcagtcaaacaggttcacgtaaattgaaacggagagagtatcatttttattaaggggtataaatatattt

1131 ................at.........ct................-----t.g..--..-......------------------------------aa..
1461 taattagttgagacttgcacatacaagtaaaatatttcttagaatacaaaatcaactgaaagcttacttctaattatatggttttgaattttcctttcaa

1194 .c..c.g...g.g..................atat......gg.c..t.......t...........a...catct.............-----......
1561 tgaagt-aaa-taaaaaggaaacaattatat----tcaacgcatgtaggtatatggtcctgtcatta-tct-----caaatcaaatggtttaaagacaaa

1289 .................................ta..----.--...--.................g........t......ta.............. ..
1649 ggactttggaaacatagaattgtcagctttatagttatggagtactatattagttagctgtttgcatctattcataattggtctatctgtgtgcag CA
 338                                                                                                    A

1379 ... ... ... ... ... ... ... ... ... ... ... ..C ... ... ... ... ... ... ... ... ... ... ... ... ...
1747 TGG TGG TAC AAG CCA GAG TAT ATC ATC AAT GAG CTT AAT ATT AAC TCT GTC ATT ACG ACG CCG TGT CAT GAA GAA
 339  W   W   Y   K   P   E   Y   I   I   N   E   L   N   I   N   S   V   I   T   T   P   C   H   E   E

1454 ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... . ............t.
1822 ATT TTG CCT ATT AAC GCC TGG ACG ACT CAG CGA CCT TAC ACG TTG AGG GGC TAT TCT TAT TCT G gttagtatttttat
 364  I   L   P   I   N   A   W   T   T   Q   R   P   Y   T   L   R   G   Y   S   Y   S

1532 tttcttctct.................a...t..............a.......aa............................................
1900 a---------ttttccgatttgctgagaatatcatatttcttagttttgtcga--tacatcgtatcctctaactctgacgttttacttcgtccttatgc

1632 ..........t..--------------------------------------.................a...........c...........--------
1989 acccacttacgtccttactttctcagacagtttattgatgaaaactacttactatttcgacccgatagcctcagcgtccttaattaaatgtgatgtttt

//// ----------------------------------------------------------------------------------------------------
2089 gaaagagatattctctcccgtctatttaattaattttggctgttttatacgtgggaatctattttaacattaattaatatagaaatgaaccatatt

//// ----------------------------------------------------------------------------------------------------
2189 aatattattaatttcttcattgaaaatacaacaaatactcttcggctcttactacaatgacaattttgaagaaaaataattaattccttcctaatatctg

1687 ------------------------------------------------------------------------................---.........
2289 aaaaatcaaatattgtggaccataaaaaaaggtcaaaaaattaattaaaatgaactggagagagtaaattagaaaatataattatagcactagtaattaa

1712 ....cg...............g.........a........t..a.................................c....................c.
2389 agttattagatgtcttctttaaaaagcgtgtgaaaactttaaagacgaaatataatatgaatattatctaatacttagaaagtgtcaataattggtagac

1812 ...............g-.--............g.......g....g....a...-.....g......c.a......g.......................
2489 aatttaaactatatactagttaaaaagtctgtcaatacaactattagtattggggattagagagaatagtagtaaaatggagtaattggacgcatgagct

1908 ....a--------...........a................c...--......................................-----.c........
2589 tgggcatgctgattgctgtcagcttgtttgctaatgtgaaaaagaaaatagtaagaaaaggccaacatggttttgtttattttattatgtggtagtacac

1993 .......................................-------..--..........c.......................................
2689 aaaaacctggggagctttcctagttctgaagagtcggtctttggtagcacaaaattaatagtatagtataccaagtgaatattaaattcaattgtctaaa

2083 ................................................g......a.....gc..g.........t..................g.....
2789 gcacggaatcttttgactactttagttcctgcatcttgggttgcctcaacaacaccctttattgaattattatagtaatgttcaatataatatacaatt

2183 ......tt..-.......................................c.....c.t................................g........
2889 agaaaacactctaagtggtcactttatatggatctagtcaatactatttcttctaaacaacgtgcctaattacttcccactttccagtacatgaccacca

2283 ........-------....................a..t..--------------------------------................c....t.....
2989 ttaagtttaattttgtcaattccttgtgcaattggcccttcaaatgagcagaagtgttacgtaggaaaactaacttcagctactattataggagtaaac

2345 a..........-----t..a..-..t...tg.....a.-................................................----....... ..
3089 ctgttaggaaaagatgctcgaggaactgac-aaaacttgtagaataattagccattgtattgattgaaatactgattgtgaacgtgtaacaaacag GC
 385                                                                                                    G

2432 ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... T.. ... ... ... ...
3186 GGA GGG AAA AAA GTA ACG CGA GTA GAA GTG ACG TTG GAT GGA GGA GAA ACA TGG CAA GTT AGC ACA CTA GAT CAC
 386  G   G   K   K   V   T   R   V   E   V   T   L   D   G   G   E   T   W   Q   V  C/S  T   L   D   H

2507 ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
3261 CCA GAG AAG CCC ACC AAA TAT GGC AAG TAC TGG TGT TGG TGC TTT TGG TCA CTC GAG GTT GAG GTG TTA GAC TTG
 411  P   E   K   P   T   K   Y   G   K   Y   W   C   W   C   F   W   S   L   E   V   E   V   L   D   L

2582 ... ... ..C ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ..T ... ... ... ... ... ..T
3336 CTC AGT GCT AAA GAA ATT GCT GTT CGA GCT TGG GAT GAG ACC CTC AAT ACT CAA CCC GAG AAG CTT ATT TGG AAC
 436  L   S   A   K   E   I   A   V   R   A   W   D   E   T   L   N   T   Q   P   E   K   L   I   W   N

2657 ... ... ...a......a...c....-......----.......g.......g....at...t....c.-................-............
3411 GTC ATG gtacgttcacttcttctttacctttatttcttttaacttctatatactagcggtgtaaagttatttacaccataagttaacttacaaaaa
 461  V   M

2748 c.....-...a....-----------------------------------------------------------------------------........
3509 tatgtaactatttatactacgagtgatgagggcaagaaggggtttaagtatttgacaataaatgtaaaccctgcaatttgttcctaatttttatcctt

2770 .....c.t...c......t............................-----------------------.c.-.t.....-............g.....
3609 tcaac-tctttgtgattgcttcattatctagattcacagagcacatgtgttcacatgccaaaacaaaaaactacaaacaaaaaactttcactagcttt

2845 .............-.....g........-.................................................-g..............g.....
3708 agtctaagattccccttttttttttgggaggtgtgtggtccatactccatagatcaattccagccactgacgtaccaaaccctgaaaattcctagtagt

2942 ......a.....................c..............-.............c.........................tt...............
3808 tatagcgacgtacaatcatttcatattatgtaagcagagacgtgatcacatgaactagatgtgaataccacttgcccagtccaccaggtcaattcatcta

3041 ...a.aca................a............g.............tag.g..........c............t......c..at........-
3908 gatgt---gtaaatcttgacaccagcactgggtcact-tttataacactag---catttaacaaca-tttcatccttgaacattacttgggctaattaat

3140 --g-........--..g...........................a..........t..t.a.......cc.............g...-------------
4000 aagtatttttttttatatactctaaaaattgtaattacataaatgaatttaacttatacacgctgacaatgttactaattccacttttacggacggtta
```

```
3223 ---------..............t..cg.............---....-.t.g-....aa.....a..........................c.......
4100 tctatagaaatcatttaggtgaaacaattctcttacactatgatcagtgttagtacataatggttat-tacattttctaaatattgtgctatgttgcaat

3309 t..... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
4199 gttcag GGA ATG ATG AAT AAT TGC TGG TTC CGA GTA AAG ATG AAT GTG TGC AAG CCT CAC AAG GGA GAG ATT GGA
 463         G   M   M   N   N   C   W   F   R   V   K   M   N   V   C   K   P   H   K   G   E   I   G

3384 ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ..A ... ... ..G ... ... ... ...
4274 ATA GTG TTT GAG CAT CCG ACT CAA CCT GGA AAC CAA TCA GGT GGA TGG ATG GCG AAG GAG AGA CAT TTG GAG ATA
 486  I   V   F   E   H   P   T   Q   P   G   N   Q   S   G   G   W   M   A   K   E   R   H   L   E   I

3459 ... ... ... ... ... .C. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
4349 TCA GCA GAG GCA CCT CAA ACA CTA AAG AAG AGT ATC TCA ACT CCA TTC ATG AAC ACA GCT TCC AAG ATG TAC TCC
 511  S   A   E   A   P  P/Q  T   L   K   K   S   I   S   T   P   F   M   N   T   A   S   K   M   Y   S

3534 ... ..G ... ..G ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
4424 ATG TCC GAG GTC AGG AAA CAC AGC TCT GCT GAC TCT GCT TGG ATC ATA GTC CAT GGT CAT ATC TAT GAC GCC ACG
 536  M   S   E   V   R   K   H   S   S   A   D   S   A   W   I   I   V   H   G   H   I   Y   D   A   T

3609 ... ... ... ... ... ... ..C ... ..T T.. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
4499 CGT TTC TTG AAA GAT CAC CCT GGT GGG ACT GAC AGC ATT CTC ATC AAT GCT GGC ACT GAT TGC ACT GAG GAA TTT
 561  R   F   L   K   D   H   P   G   G  S/T  D   S   I   L   I   N   A   G   T   D   C   T   E   E   F

3684 ... ... ... ... ... ... ... ... ... ... ..A ... ... ..A ... ... ... ... ... ... C.. ... ... ... ...
4574 GAT GCA ATT CAT TCT GAT AAG GCT AAG AAG CTC TTG GAG GAT TTC AGG ATT GGT GAA CTC ATA ACT ACT GGT TAC
 586  D   A   I   H   S   D   K   A   K   K   L   L   E  E/D  F   R   I   G   E   L  L/I  T   T   G   Y

3759 ... ... ... ... ... ... ... ... ..C ..T ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
4649 ACC TCT GAC TCT CCT GGC AAC TCC GTG CAC GGA TCT TCT TCC TTC AGC AGC TTT CTA GCA CCT ATT AAG GAA CTT
 611  T   S   D   S   P   G   N   S   V   H   G   S   S   S   F   S   S   F   L   A   P   I   K   E   L

3834 ... ... ... ... ... ... ... ... ..C ... ... ... ... ... ... ... ... ... ... ... ... ..A ... ... ...
4724 GTT CCA GCG CAG AGG AGT GTG GCC CTA ATT CCA AGA GAG AAA ATC CCA TGC AAA CTC ATC GAC AAG CAA TCC ATC
 636  V   P   A   Q   R   S   V   A   L   I   P   R   E   K   I   P   C   K   L   I   D   K   Q   S   I

3909 ... .C. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
4799 TCC CAT GAT GTT AGG AAA TTT CGA TTT GCA TTG CCC TCT GAG GAT CAA GTC TTG GGC TTG CCT GTT GGA AAA CAT
 661  S  P/H  D   V   R   K   F   R   F   A   L   P   S   E   D   Q   V   L   G   L   P   V   G   K   H

3984 ... ... ... ... ... ... ... ... ... ... ... ... ... ... ..C ... ... ... ... ... ... ... ... ... ...
4874 ATC TTC CTC TGT GCC GTT ATT GAC GAT AAG CTC TGC ATG CGC GCT TAC ACG CCT ACT AGC ACG ATC GAT GAG GTG
 686  I   F   L   C   A   V   I   D   D   K   L   C   M   R   A   Y   T   P   T   S   T   I   D   E   V

4059 ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ..G ... ... ... ...
4949 GGG TAC TTC GAG TTG GTT GTC AAG ATA TAC TTC AAA GGA ATT CAC CCT AAA TTC CCC AAT GGA GGG CAA ATG TCA
 711  G   Y   F   E   L   V   V   K   I   Y   F   K   G   I   H   P   K   F   P   N   G   G   Q   M   S

4134 ..A ..C ... ... ... C.C .AA ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
5024 CAG TAT CTT GAT TCT ATG CCG TTA GGG TCA TTT CTC GAC GTG AAA GGT CCA TTA GGT CAC ATT GAA TAC CAA GGA
 736  Q   Y   L   D   S  L/M Q/P  L   G   S   F   L   D   V   K   G   P   L   G   H   I   E   Y   Q   G

4209 ... ..C ... ... ... ... ... ... ... ..A ... ... ... ... ... ... ... ... ... ... ... ... ... ..G ...
5099 AAG GGA AAT TTC TTA GTT CAT GGC AAA CAG AAG TTT GCC AAG AAG TTG GCC ATG ATA GCA GGT GGA ACA GGA ATA
 761  K   G   N   F   L   V   H   G   K   Q   K   F   A   K   K   L   A   M   I   A   G   G   T   G   I

4284 ... ... ..T ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ..C ... ...
5174 ACT CCA GTG TAT CAA GTC ATG CAG GCA ATT CTG AAA GAT CCA GAA GAT GAC ACA GAA ATG TAT GTG GTG TAT GCT
 786  T   P   V   Y   Q   V   M   Q   A   I   L   K   D   P   E   D   D   T   E   M   Y   V   V   Y   A

4359 ..T ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ..A ... ... ...
5249 AAC AGA ACA GAG GAT GAT ATT TTA CTT AAG GAA GAG CTT GAT TCA TGG GCT GAG AAA ATT CCA GAG AGG GTT AAA
 811  N   R   T   E   D   D   I   L   L   K   E   E   L   D   S   W   A   E   K   I   P   E   R   V   K

4434 ... ... ... ... ... ..A ... ... ... ... ... ... ... ... ... ... C.. ... ... ... T.. ... ... ... ...
5324 GTT TGG TAT GTG GTT CAG GAT TCT ATT AAA GAA GGA TGG AAG TAC AGC ATT GGT TTT ATT ACA GAA GCC ATT TTG
 836  V   W   Y   V   V   Q   D   S   I   K   E   G   W   K   Y   S  L/I  G   F   I  S/T  E   A   I   L

4509 ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
5399 AGA GAA CAT ATC CCT GAG CCA TCT CAC ACA ACA CTG GCT TTG GCT TGT GGA CCA CCT CCT ATG ATT CAA TTT GCT
 861  R   E   H   I   P   E   P   S   H   T   T   L   A   L   A   C   G   P   P   P   M   I   Q   F   A

4584 ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... *** ....c..............
5474 GTT AAT CCA AAC TTG GAG AAG ATG GGC TAT GAC ATT AAG GAT TCC TTA TTG GTG TTC TAA ttttaaaacaaaacaatat
 886  V   N   P   N   L   E   K   M   G   Y   D   I   K   D   S   L   L   V   F  904

4664 ........****.c.....g.......-...........................a..................c.c.a.................c.tc
5554 ctgcaggaataaatttttttttcccccctatcagttgtacatattgtatttggtttatcaccccatgtactacgtagtgtttgtagttcttacattt--

4763 ..c..c...tt......ttt...............****......aattt........t.........g..................--...........
5652 ttattttttagaatttt---tttaaaccttaggatataaaggttt-----tctcttccaacaaagtgattctttagggaagaaatgtactgtactgtact

4861 .........g..t.aa..g.........................................g.......................................
5744 agtatgtctaag-ccgaaagttgtaatgtttaccatgacaaattgtattcaattcctcatggaatagtaacattgtgttcatgtgtcttcctgtaagcga

4961 .t...............gtgtgt...........tatatatatatatatatatatatatatatatatatatatatatata..........cggt........
5843 t-cttcaaaatatcaat------gtatatatata--------------------------------------------gtaattgcaaaccattgttcct

5061 ......t....g.......g.....a.c...............................c...ctttt....a..t.----.............****..
5892 tttcccgatgtagttaactactctttctttagcttctagtctctggtgaatattttttttt----tctataactctttaattaatacggccttaaataaga

5157 .......a............................-...t..........................----.........t...................
5988 gaaaagtttaaaccacgaatatcattatgcagacgtata-ggtaattaatctactttttgaaaaaaaatctattttctttatgtggtccttcaaaataat

5252 ..............................................  5297          N.tomentosiformis genome    Nia-1
6087 attctagaaccttttgtatattcccttttaacttctatttagtttt  6132          N.sylvestris genome         Nia-2
```