# Primer Sequences for coxiii (mal\_mito\_1), species detection

RNA10 Single Exon (M76611: 625-724, REVERSE, 100 nt)

Region identified by D. Echeverry (M76611: 631-1138, 508 nt)

COX3 Single Exon (M76611: 734-1486, REVERSE, 753 nt)

60

1234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890

TTTGATAGCGGTTAACCTTTCCTTTTCCTTACGTACTCTAGCTATGAACACAATTGTCTATTCGTACAATTATTCA

TATATATATTTGAAACAGGACATAcatgttcatTTATTCTGAATAGAATAAGAACTCTATAAATAACCAGATTATTTCAACAAAATGCCAATATAAAATT

GTAATTTGATCAGTGTGAGGTATAACAATATATGATATACCGAAAGAATTTATAAACCATTCGGTAGAAGTATCATATATTTCTATTATTCTTATAAAGT

ATATTATTAATAATAATAAACCTATTACTACATGAGAAAAATGTAATCCTGTAACACAATAAAATAATGTAGTATATACAGTATCATTTATATGATATGA

TAAATGTAAATACTCTGTAGTTTGTAGAGATGCAAAACATTCTCCTAATAAGTATATTATACAAATAATACTAGAGATTTCAAAACTCATTCCTTTTTCT

ATAAATACTTGTAAACATGCAGTCATACATGATGCACTAGCTAATATAAATGTAATTGTTAAGATTAACATTCTTGATGAAGTAATGATAATACCTTCAT

TACTTAATGGATATGGTGATAAACTAAAATGTAATATACCCCAAAAATATGTAAAGAATAATAAAGCTTCTGATATTATGATAGATAACATACCAGAAGT

TAAAGATGAAAATACAGAATAAAAACTTTCTCGAATAGAATATACAAATATTAATAGGATTATAGGGTTAAATGTAAATAATATCCCTACAGAAAAGTAT

TTTAAAGATGTACCATATAATGATGTTAATGCAGGATATGAAACTAGATGTGCTTTTATATTTGATAAATTACTAAATAAAATAAA

## First Stage of analysis

One primer, two targets

6 7

2 3 4 5 6 7 8 9 0 1

01234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901 23456789

Common CGCCATTTGATAGCGGTTAACCTTTC-TTT-CCTTACGTACTCTAGCT----ACACAA------------------------------------GAAACAGG

Pf --------------------------C---T-----------------ATGA------TTGTCTATTCGTACAATTATTCATATATATATTT----------

Pf[2] --------------------------C---T-----------------ATGA------TTGTCTATTCGTACAATTATTCATATATATATATTT--------

Pm --------------------------C---T-----------------TTGT------ATTAATTCGTCTACATTAGATACTATATTT--------------

Pow --------------------------C---T-----------------ATTT------ACATAATATCGTCAAATATAAGTACTTTATTC------------

Poc --------------------------C---T-----------------ATTT------ATATATTATTGTCAAATATAAGTACTTTAATC------------

Pv --------------------------T---T-----------------TTTA------TATTATTGTCTATACTAGATACTATAGTT---------------

Pv[2] --------------------------T---T-----------------TTTA------TATTATTGTCTATTACTAGATACTATAGTT--------------

Pk --------------------------T---------------------TTTT------TATTATTGTCTATACTAGATACTATGGAC---------------

7 8

2 3 4 5 6 7 8 9 0 1

01234567890123---45678901234567890123456789012345678901234567890123456789012345678901234567890123456789

Common ACATA------------TTATTC-GAATAGAA---GAACTCTATAAATAAC

Pf -----CATGTTCAT---------T--------TAA----------------

Pf[2] -----CATGTTCAT---------T--------TAA----------------

Pm -----TATGTTCAA---------T--------TAG----------------

Pow -----TATGTTACA---------A--------TAA----------------

Poc -----TATGTTACA---------A--------TAA----------------

Pv -----TACATATATTCA------T--------AAA----------------

Pv[2] -----TACATATATTCA------T--------AAA----------------

Pk -----TATTCA------------T--------AAA----------------

**Locus Anchor: 651** CCTTACGTACTCTAGCT....ACACAA **Target: 668-671 Target: 678-691**

Pf ATGA Pf TTGTCTATTCGTAC

Pm TTGT Pm ATTAATTCGTCTAC

Pow/Poc ATTT Pow ACATAATATCGTCA

Poc ATATATTATTGTCA

Pv TTTA Pv TATTATTGTCTATA

Pv TATTATTGTCTATT

Pk TTTT Pk TATTATTGTCTATA

**Config**

species.loci=mito1

species.locus.mito1.region=M76611:520-820

species.locus.mito1.anchors=651@CCTTACGTACTCTAGCT....ACACAA

species.locus.mito1.targets=species1@668-671,species2@678-691

species.locus.mito1.target.species1.alleles=Pf@ATGA,Pv@TTTA,Pm@TTGT,Pow|Poc@ATTT,Pk@TTTT

species.locus.mito1.target.species2.alleles=Pf@TTGTCTATTCGTAC,Pv@TATTATTGTCTATT|TATTATTGTCTATA,Pm@ATTAATTCGTCTAC,Pow@ACATAATATCGTCA,Poc@ATATATTATTGTCA,Pk@TATTATTGTCTATA

**Analysis of common unclassified alleles**

i.e. found in our sample set, and unlikely to be only sequencing errors.

We analyzed all alleles that were found at target position, but were not listed in the configuration file. We then discarded:

* All singleton reads in any one sample
* All alleles at <1% of the reads in a sample
* All alleles with a single nucleotide different, and at <2% of the reads in a sample
* All alleles at very low read count (<5) in samples with v. low read count (<10)

The following unclassified alleles were found to be important:

|  |  |  |  |
| --- | --- | --- | --- |
| Allele | Samples | Species | Diff |
| ACATAATATTGTCA | 6 | Pow | 1 |
| TTGTCTAATCGTAC | 6 | Pf | 1 |
| TTGTCTACTCGTAC | 4 | Pf | 1 |
| TTGTCTATTCATAC | 3 | Pf | 1 |
| TTGTCTATTTGTAC | 3 | Pf | 1 |
| TTGTCTGTTCGTAC | 19 | Pf | 1 |
| TTGTTTATTCGTAC | 2 | Pf | 1 |
| TTGTTTCATTTGAT | 8 | Pf | 7 |

When looking in detail, we found the differences to be the following:

Pow[2] ACATAATATTGTCA

Pf[2] TTGTCTAATCGTAC

Pf[3] TTGTCTACTCGTAC

Pf[4] TTGTCTATTCATAC

Pf[5] TTGTCTATTTGTAC

Pf[6] TTGTCTGTTCGTAC

Pf[7] TTGTTTATTCGTAC

Pf[8] TTGTTTCATTTGAT This one is odd: not found in BLAST. Do not add for now, investigate further

This suggests that considerable variability occurs beyond position 6, suggesting that the allele analysis should be restricted to positions 1-6 (or rather, 678-683): as follows

**Locus: mito1** M76611:520-820  
**Anchor: 651** CCTTACGTACTCTAGCT....ACACAA **Target: 668-671 Target: 678-683**

Pf ATGA Pf TTGTCT

TTGTTT

Pm TTGT Pm ATTAAT

Pow/Poc ATTT Pow ACATAA

Poc ATATAT

Pv TTTA Pv/Pk TATTAT

Pk TTTT

Ideally, we should now find some target that can confirm Pv vs. Pk and Pow vs Poc

## Second Stage of analysis

More primers and targets

Identifying more targets that can separate species

**Locus mito1**

62 3 4 5 6 7 8 9 70 1

01234567890123456789012345678901234567890123456789012345678901234567890123456789012345678--90123456789

Common CGCCATTTGATAGCGGTTAACCTTTC-TTT-CCTTACGTACTCTAGCT----ACACAA------------------------------------GAAACAGG

Pf --------------------------C---T-----------------ATGA------TTGTCTATTCGTACAATTATTCATATATATA--TTT--------

Pf[2] TA

Pm --------------------------C---T-----------------TTGT------ATTAATTCGTCTACATTAGATACTATATTT--------------

Pow --------------------------C---T-----------------ATTT------ACATAATATCGTCAAATATAAGTACTTTATTC------------

Poc --------------------------C---T-----------------ATTT------ATATATTATTGTCAAATATAAGTACTTTAATC------------

Pv --------------------------T---T-----------------TTTA------TATTATTGTCTA-TACTAGATACTATAGTT--------------

Pv[2] T

Pk --------------------------T---------------------TTTT------TATTATTGTCTA-TACTAGATACTATGGAC--------------

**Locus: mito1** M76611:520-820 **Anchor: 651** CCTTACGTACTCTAGCT....ACACAA **Target: 668-671 Target: 678-683**

Pf ATGA Pf TTGTCT

TTGTTT

Pm TTGT Pm ATTAAT

Pow/Poc ATTT Pow ACATAA

Poc ATATAT

Pv TTTA Pv/Pk TATTAT

Pk TTTT

**Config**

species.locus.mito1.region=M76611:520-820

species.locus.mito1.anchors=651@CCTTACGTACTCTAGCT....ACACAA

species.locus.mito1.targets=species1@668-671,species2@678-683

species.locus.mito1.target.species1.alleles=Pf@ATGA,Pv@TTTA,Pm@TTGT,Pow|Poc@ATTT,Pk@TTTT

species.locus.mito1.target.species2.alleles=Pf@TTGTCT|TTGTTT,Pv|Pk@TATTAT,Pm@ATTAAT,Pow@ACATAA,Poc@ATATAT

**Locus mito2**

72 3 4 5 6 7 8 9 80 1

01234567890123---45678901234567890123456789012345678901234567890123456789012345678901234567890123456789

Common ACATA------------TTATTC-GAATAGAA---GAACTCTATAAATAACCA----ATTTC---AAAATGCCA-TATAA-ATTGT-ATTTGATC----TGAG

Pf -----CATGTTCAT---------T--------TAA------------------GATT-----AAC---------A-----A-----A--------AGTG----

Pf GACT

Pm -----TATGTTCAA---------T--------TAG------------------TACT-----TAT---------A-----T-----T--------AGTA----

Pow -----TATGTTACA---------A--------TAA------------------TATT-----TAC---------A-----T-----G--------TGTG----

Poc -----TATGTTACA---------A--------TAA------------------TATT-----TAT---------A-----T-----T--------AGTA----

Pv -----TACATATATTCA------T--------AAA------------------TACA-----AAC---------G-----T-----A--------AGTA----

Pk -----TATTCA------------T--------AAA------------------TACA-----AAC---------G-----T-----A--------AGTA----

**Locus: mito2** M76611:620-920 **Anchor: 752** GAACTCTATAAATAACCA....ATTTC **Target: 770-781**

Pf GATTATTTCAAC

GACTATTTCAAC

Pm TACTATTTCTAT

Pow TATTATTTCTAC

Poc TATTATTTCTAT

Pv/Pk TACAATTTCAAC

**Config**

species.locus.mito2.region=M76611:620-920

species.locus.mito2.anchors=752@GAACTCTATAAATAACCA....ATTTC

species.locus.mito2.targets=species3@770-781

species.locus.mito2.target.species3.alleles=Pf@GATTATTTCAAC|GACTATTTCAAC,Pv|Pk@TACAATTTCAAC,Pm@TACTATTTCTAT,Pow@TATTATTTCTAC,Poc@TATTATTTCTAT

**Locus mito3**

82 3 4 5 6 7 8 9 90 1

0123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890123456789

Common GTA-------ATA-GATATACC--------------ACCAT------------TC-TA----TC----ATTCTTATAAA-TATAT---TAATAA-AATAA

Pf ---TAACAAT---T--------GAAAGAATTTATAA-----TCGGTAGAAGTA--A--TATT--TATT-----------G-----TAT------T-----

Pm ---AAATAAC---T--------ATATAATGAATATG-----TCACTATTTATG--A--AATA--TACT-----------G-----TAA------T-----

Pow ---ATATAAC---T--------ATATAATGAATATG-----CCCTTATTTGTA--A--AACT--AACC-----------G-----CAT------A-----

Poc ---AAATAAC---T--------ATATA**G**TGAATATG-----TCGTTATTAGTA--A--AACT--AACC-----------G-----CAT------A-----

Pv ---ATACAAT---A--------ATATAATGAATATG-----TCACTATTAGTA--A--CATT--AACT-----------A-----TAA------T-----

Pk ---ATAATAT---A--------ATATAATGAATATG-----TCACAATTAGTA--G--AACT--TATT-----------A-----TAA------T-----

**Locus: mito3** M76611:720-1020 **Anchor: 842** (?:GAAAGAATTTATAA|ATATA[AG]TGAATATG)ACCAT

**Target: 861-869 Target: 878-887**

Pf TCGGTAGAA Pf TATTTCTATT

Pm TCACTATTT Pm AATATCTACT

Pow CCCTTATTT Pow/Poc AACTTCAACC

Poc TCGTTATTA

Pv TCACTATTA Pv CATTTCAACT

Pk TCACAATTA Pk AACTTCTATT

**Config**

species.locus.mito3.region=M76611:720-1020

species.locus.mito3.anchors=842@(?:GAAAGAATTTATAA|ATATA[AG]TGAATATG)ACCAT

species.locus.mito3.targets=species4@861-869,species5@878-887

species.locus.mito3.target.species4.alleles=Pf@TCGGTAGAA,Pv@TCACTATTA,Pm@TCACTATTT,Pow@CCCTTATTT,Poc@TCGTTATTA,Pk@TCACAATTA

species.locus.mito3.target.species5.alleles=Pf@TATTTCTATT,Pv@CATTTCAACT,Pm@AATATCTACT,Pow|Poc@AACTTCAACC,Pk@AACTTCTATT

**Locus mito4**

92 3 4 5 6 7 8 9 100 1

0123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890123456789

Common ACC-------ACATG-GAAAAATGTAA-CCTGTAACACAATAAAATAATGT------------ATCATT-ATA--ATAT—TAAATGTAAA-TA-TC-GT-

Pf ---TATTACT-----A-----------T-----------------------AGTATATACAGT------T---TG----GA---------A--C--T--A

Pm ---AATAATA-----T-----------A-----------------------TCCTGAAACTCC------A---TA----TC---------A--T--A--T

Pow ---AACAATA-----T-----------A-----------------------ACCAGATATAGC------A---TA----CC---------A--T--A--A

Poc ---AACAATA-----T-----------A-----------------------ACCAGATATAGC------A---TA----CC---------A--T--A--A

Pv ---AACTATA-----T-----------A-----------------------ACCAGATATAGC------T---CA----CC---------G--T--A--T

Pk ---AACTATA-----T-----------A-----------------------ACCTGATATAGC------T---CA----CC---------G--T--A--T

**Locus: mito4** M76611:820-1120 **Anchor: 948** CCTGTAACACAATAAAATAATGT

**Target: 971-973**

Pf AGTA

Pm TCCT

Pow/Poc/Pv ACCA

Pk ACCT

**Config**

species.locus.mito4.region=M76611:820-1120

species.locus.mito4.anchors=948@CCTGTAACACAATAAAATAATGT

species.locus.mito4.targets=species6@971-974

species.locus.mito4.target.species6.alleles=Pf@AGTA,Pv|Pow|Poc@ACCA,Pm@TCCT,Pk@ACCT

**Locus mito5/mito6**

102 3 4 5 6 7 8 9 110 1

0123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890123456789

Common GT-TG----GATGCAAAACATTCTCC----AA-TA-ATTATA-A-A-AATACT-GA-ATTTC-AAACTCATTCCTTTTTCTA-------TTGTAAACATG

Pf --T--TAGA-----------------TAAT--G--T------C-A-T------A--G-----A--------T----------TAAATAC-----------

Pm --C--TAAT-----------------AAGT--A--T------A-T-T------T--T-----A--------A----------AAAGAAA-----------

Pow --T--TAAT-----------------AAGA--A--T------A-T-T------T--T-----T--------T----------TGAGAAA-----------

Poc --T--TAAT-----------------AAGG--A--A------A-T-T------T--T-----T--------T----------TAAGAAA-----------

Pv --T--AAGT-----------------AAGT--G--T------A-T-C------A--A-----T--------T----------TTAGAAA-----------

Pk --T--AAGT-----------------AAGT--G--T------A-T-T------A--A-----T--------T----------TTAGAAA-----------

**Locus: mito5** M76611:900-1200 **Anchor: 1029** GATGCAAAACATTCTCC  
**Target: 1046-1049**

Pf TAAT

Pm/Pv/Pk AAGT

Pow AAGA

Poc AAGG

**Config**species.locus.mito5.region=M76611:900-1200

species.locus.mito5.anchors=1029@GATGCAAAACATTCTCC

species.locus.mito5.targets=species7@1046-1049

species.locus.mito5.target.species7.alleles=Pf@TAAT,Pv|Pm|Pk@AAGT,Pow@AAGA,Poc@AAGG

**Locus: mito6** M76611:950-1250 **Anchor: 1077** ATTTC[AT]AAACTCATTCCTTTTTCTA  
**Target: 1062-1066 Target: 1102-1105**

Pf CAAAT Pf TAAA

Pm/Pow/Poc/Pk AATAT Pm AAAG

Pow TGAG

Poc TAAG

Pv AATAC Pv/Pk TTAG

**Config**species.locus.mito6.region=M76611:950-1250

species.locus.mito6.anchors=1077@ATTTC[AT]AAACTCATTCCTTTTTCTA

species.locus.mito6.targets=species8@1062-1066,species9@1102-1105

species.locus.mito6.target.species8.alleles=Pf@CAAAT,Pv@AATAC,Pm|Pow|Poc|Pk@AATAT

species.locus.mito6.target.species9.alleles=Pf@TAAA,Pv|Pk@TTAG,Pm@AAAG,Pow@TGAG,Poc@TAAG

#### Full Config

species.classes=Pf,Pv,Pm,Pow,Poc,Pk

species.loci=mito1,mito2,mito3,mito4,mito5,mito6

species.locus.mito1.region=M76611:520-820

species.locus.mito1.anchors=651@CCTTACGTACTCTAGCT....ACACAA

species.locus.mito1.targets=species1@668-671,species2@678-683

species.locus.mito1.target.species1.alleles=Pf@ATGA,Pv@TTTA,Pm@TTGT,Pow|Poc@ATTT,Pk@TTTT

species.locus.mito1.target.species2.alleles=Pf@TTGTCT|TTGTTT,Pv|Pk@TATTAT,Pm@ATTAAT,Pow@ACATAA,Poc@ATATAT

species.locus.mito2.region=M76611:620-920

species.locus.mito2.anchors=752@GAACTCTATAAATAACCA....ATTTC

species.locus.mito2.targets=species3@770-781

species.locus.mito2.target.species3.alleles=Pf@GATTATTTCAAC|GACTATTTCAAC,Pv|Pk@TACAATTTCAAC,Pm@TACTATTTCTAT,Pow@TATTATTTCTAC,Poc@TATTATTTCTAT

species.locus.mito3.region=M76611:720-1020

species.locus.mito3.anchors=842@(?:GAAAGAATTTATAA|ATATA[AG]TGAATATG)ACCAT

species.locus.mito3.targets=species4@861-869,species5@878-887

species.locus.mito3.target.species4.alleles=Pf@TCGGTAGAA,Pv@TCACTATTA,Pm@TCACTATTT,Pow@CCCTTATTT,Poc@TCGTTATTA,Pk@TCACAATTA

species.locus.mito3.target.species5.alleles=Pf@TATTTCTATT,Pv@CATTTCAACT,Pm@AATATCTACT,Pow|Poc@AACTTCAACC,Pk@AACTTCTATT

species.locus.mito4.region=M76611:820-1120

species.locus.mito4.anchors=948@CCTGTAACACAATAAAATAATGT

species.locus.mito4.targets=species6@971-974

species.locus.mito4.target.species6.alleles=Pf@AGTA,Pv|Pow|Poc@ACCA,Pm@TCCT,Pk@ACCT

species.locus.mito5.region=M76611:900-1200

species.locus.mito5.anchors=1029@GATGCAAAACATTCTCC

species.locus.mito5.targets=species7@1046-1049

species.locus.mito5.target.species7.alleles=Pf@TAAT,Pv|Pm|Pk@AAGT,Pow@AAGA,Poc@AAGG

species.locus.mito6.region=M76611:950-1250

species.locus.mito6.anchors=1077@ATTTC[AT]AAACTCATTCCTTTTTCTA

species.locus.mito6.targets=species8@1062-1066,species9@1102-1105

species.locus.mito6.target.species8.alleles=Pf@CAAAT,Pv@AATAC,Pm|Pow|Poc|Pk@AATAT

species.locus.mito6.target.species9.alleles=Pf@TAAA,Pv|Pk@TTAG,Pm@AAAG,Pow@TGAG,Poc@TAAG

TODO

1. List of classes in configuration
2. Consensus calling
3. Same read associated to multiple loci