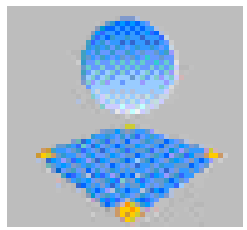
Minisatellite Markers Reveals Frequent Genetic Exchanges among House Mouse Subspecies

Eric Rivals

rivals@lirmm.fr

LIRMM, CNRS, Université Montpellier 2

http://www.lirmm.fr/~rivals



Outline

Minisatellites (MS)

- **Tandem repeat** loci present in genome of all kingdom

Minisatellites (MS)

- **Tandem repeat** loci present in genome of all kingdom
- $>$ 5% in bacteria *E. ruminatum*

Minisatellites (MS)

- **Tandem repeat** loci present in genome of all kingdom
- $>$ 5% in bacteria *E. ruminatum*
- pattern between 7 and 100 bps, up to 20 Kb long

Minisatellites (MS)

- **Tandem repeat** loci present in genome of all kingdom
- $>$ 5% in bacteria *E. ruminatum*
- pattern between 7 and 100 bps, up to 20 Kb long
- **Polymorphism**: Hypervariable in *H.s.*, ($>$ 0.5% per gamete)

Minisatellites (MS)

- **Tandem repeat** loci present in genome of all kingdom
- $>$ 5% in bacteria *E. ruminatum*
- pattern between 7 and 100 bps, up to 20 Kb long
- **Polymorphism**: Hypervariable in *H.s.*, ($>$ 0.5% per gamete)
  Simply variable in *M.m.*(add $>$ 3 units: $< 5.10^{-6}$ per gamete)

Minisatellites (MS)

- **Tandem repeat** loci present in genome of all kingdom
- $>$ 5% in bacteria *E. ruminatum*
- pattern between 7 and 100 bps, up to 20 Kb long
- **Polymorphism**: Hypervariable in *H.s.*, ($>$ 0.5% per gamete)
  Simply variable in *M.m.*(add $>$ 3 units: $< 5.10^{-6}$ per gamete)
- Mouse: simple intra-allelic tandem duplication and contraction
  [Bois et al., 2002]

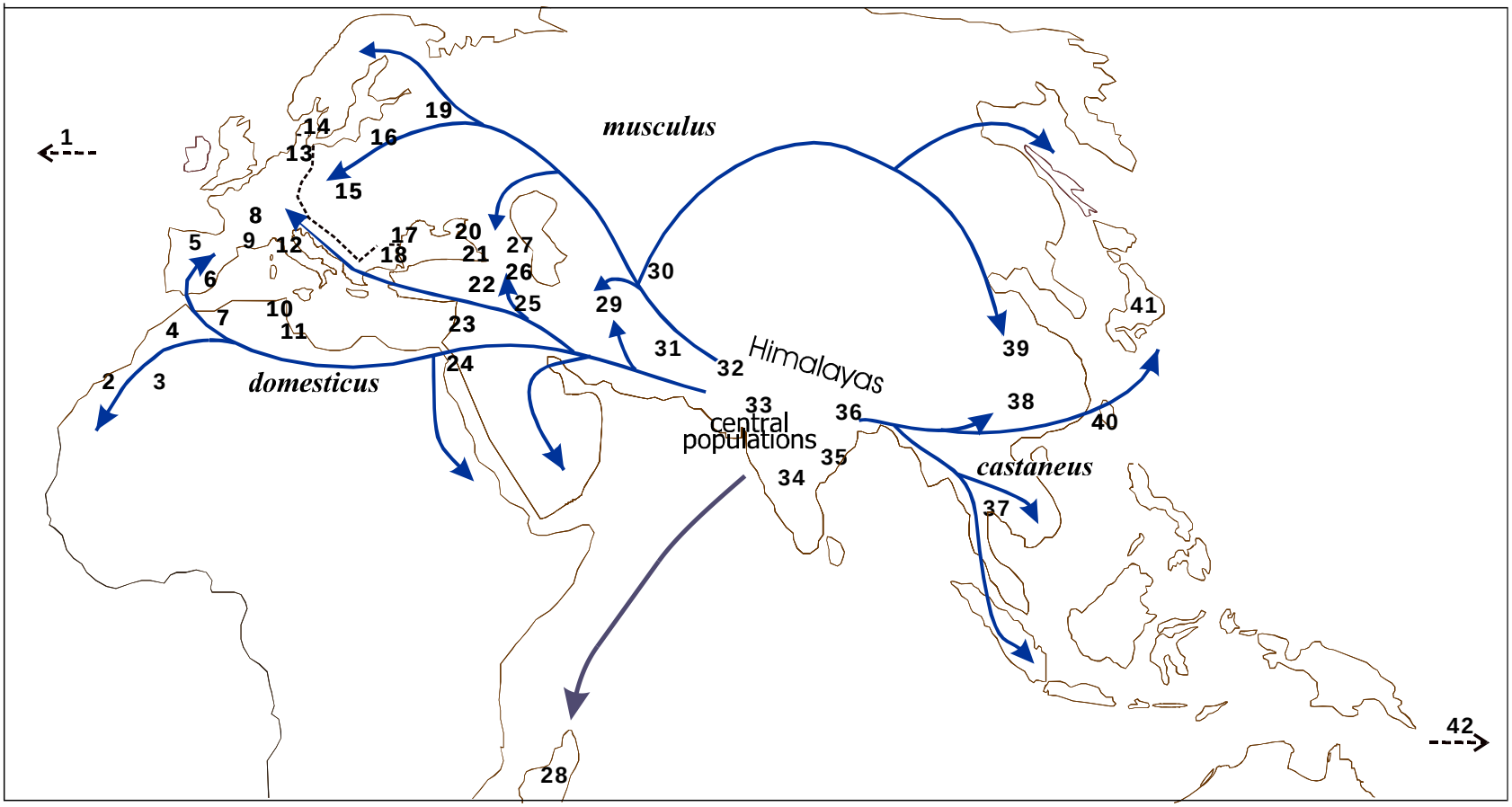
Minisatellites (MS)

- **Tandem repeat** loci present in genome of all kingdom
- $> 5\%$ in bacteria *E. ruminatum*
- pattern between 7 and 100 bps, up to 20 Kb long
- **Polymorphism**: Hypervariable in *H.s.*, ($> 0.5\%$ per gamete)
  Simply variable in *M.m.*(add > 3 units: $< 5.10^{-6}$ per gamete)
- Mouse: simple intra-allelic tandem duplication and contraction
  [Bois et al., 2002]
- MS informative for intra-species evolution

Minisatellite data

- Four minisatellite loci: MMS 24, 26, 80, and 30 respectively on chromosomes 7 (22 cM), 9 (68 cM and 79 cM), and X (43 cM)
- Panel of 116 individuals of various geographical origins
- Maps obtained by MVR-PCR as in [Bois et al., 2002]
- High diversity in length and array structure haplotypic diversity (He) in $[0.90, 0.99]$

Geographical origin of wild mice



Mouse minisatellite maps of MMS30 (X chr)

Repeat unit: 39 bps

Variant code		sequence
K	=	"aggagattcagttcacaCtatacagaagatggtgtcagc"
L	=	"aggagattcacttcagaGtatacagaagatggtgtcagc"

Mouse minisatellite maps of MMS30 (X chr)

Repeat unit: 39 bps

Variant code		sequence
K	=	"aggagattcagttcacaCtatacagaagatggtgtcagc"
L	=	"aggagattcacttcagaGtatacagaagatggtgtcagc"

Example of **maps** in wild mice

ID (species, location)	map
SPR_FRAN_Ardeche	KGKGKLKHLKKLLKYKG
CEN_INDE_Gauhati	GKKKKWGKKYKWKGWGHoGoKWKKKoLYY
CEN_INDE_Varanasi	GKKKKWGKKYKWKGWGHoGoKWKKKoKYY

Mouse minisatellite maps of MMS30 (X chr)

Repeat unit: 39 bps

Variant code		sequence
K	=	"aggagattcagttcacaCtatacagaagatggtgtcagc"
L	=	"aggagattcacttcagaGtatacagaagatggtgtcagc"

Example of **maps** in wild mice

ID (species, location)	map
SPR_FRAN_Ardeche	KGKGKLKHLKKLLKYKG
CEN_INDE_Gauhati	GKKKKWGKKYKWKGWGHoGoKWKKKoLYY
CEN_INDE_Varanasi	GKKKKWGKKYKWKGWGHoGoKWKKKoKYY
DOM_GEOR_Adjarie	GYKKKWGKLKYoWKGWGKoGGYWYKKoKYYYKG
DOM_OCEA_Tahiti	GYKKKWGKLKYoWKGWGKoGGYWYKKoKKKYYYKG
MUS_GEOR_Tbilissi	GYYKGYKYKGYKKKWGKoKYoWKYYKG

Methods

Protocol of the analysis

Input: set of sequences (maps)

Protocol of the analysis

Input: set of sequences (maps)

1. Comparison all against all $\Rightarrow$ pairwise distance matrix

Protocol of the analysis

Input: set of sequences (maps)

1. Comparison all against all $\Rightarrow$ pairwise distance matrix
   MS_Align [Bérard, Rivals, 2003]

Protocol of the analysis

Input: set of sequences (maps)

1. Comparison all against all $\Rightarrow$ pairwise distance matrix
   MS_Align [Bérard, Rivals, 2003]

2. Inference of evolutionary tree from the distances

Protocol of the analysis

Input: set of sequences (maps)

1. Comparison all against all $\Rightarrow$ pairwise distance matrix
   MS_Align [Bérard, Rivals, 2003]

2. Inference of evolutionary tree from the distances
   FastME [Desper, Gascuel, 2002]

Protocol of the analysis

Input: set of sequences (maps)

1. Comparison all against all $\Rightarrow$ pairwise distance matrix
   MS_Align [Bérard, Rivals, 2003]

2. Inference of evolutionary tree from the distances
   FastME [Desper, Gascuel, 2002]

3. Test robustness of the trees w.r.t. alignment parameters

Protocol of the analysis

Input: set of sequences (maps)

1. Comparison all against all $\Rightarrow$ pairwise distance matrix
   MS_Align [Bérard, Rivals, 2003]

2. Inference of evolutionary tree from the distances
   FastME [Desper, Gascuel, 2002]

3. Test robustness of the trees w.r.t. alignment parameters
   Criterion: Percentage of explained variance (VAF)

Protocol of the analysis

Input: set of sequences (maps)

1. Comparison all against all $\Rightarrow$ pairwise distance matrix
   MS_Align [Bérard, Rivals, 2003]

2. Inference of evolutionary tree from the distances
   FastME [Desper, Gascuel, 2002]

3. Test robustness of the trees w.r.t. alignment parameters
   Criterion: Percentage of explained variance (VAF)

4. Assess confidence of the tree internal nodes

Protocol of the analysis

Input: set of sequences (maps)

1. Comparison all against all $\Rightarrow$ pairwise distance matrix
   MS_Align [Bérard, Rivals, 2003]

2. Inference of evolutionary tree from the distances
   FastME [Desper, Gascuel, 2002]

3. Test robustness of the trees w.r.t. alignment parameters
   Criterion: Percentage of explained variance (VAF)

4. Assess confidence of the tree internal nodes
   Criterion: Rate of elementary well designed quartets (Re)
   Qualitree [Garreta, Guénoche, 2000]

Evolutionary model

- Substitution : WGY $\longrightarrow$ WKY

Evolutionary model

- Substitution : WGY $\longrightarrow$ WKY
- Deletion: WGY $\longrightarrow$ WY

Evolutionary model

- Substitution : WGY $\longrightarrow$ WKY
- Deletion: WGY $\longrightarrow$ WY
- Insertion (dual): WY $\longrightarrow$ WGY

Evolutionary model

- Substitution : WGY $\longrightarrow$ WKY
- Deletion: WGY $\longrightarrow$ WY
- Insertion (dual): WY $\longrightarrow$ WGY
- Tandem duplication: WKY $\longrightarrow$ WKKY

Evolutionary model

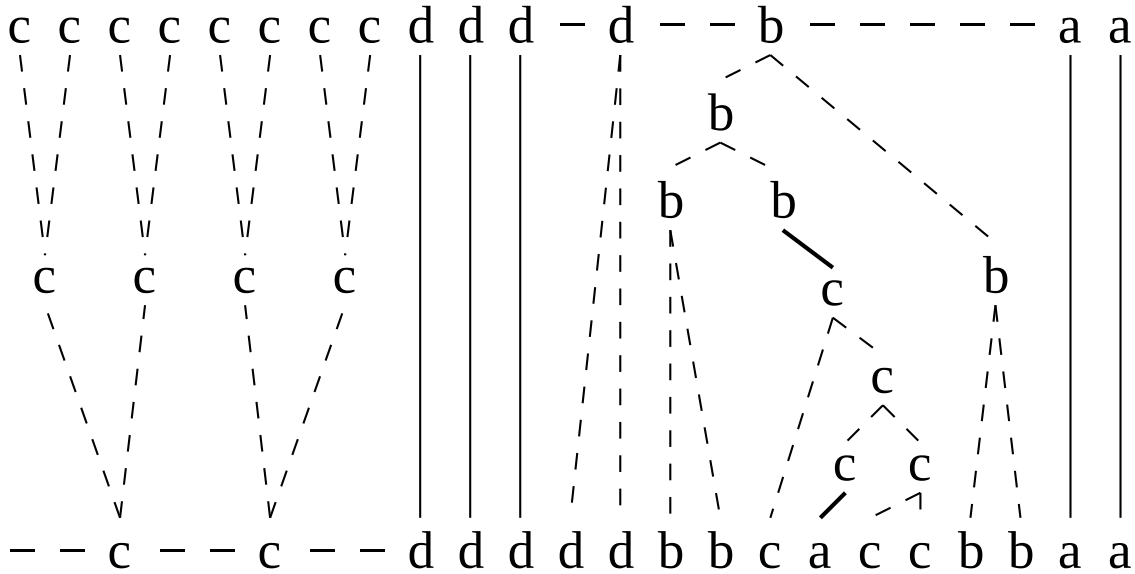
- Substitution : WGY $\longrightarrow$ WKY
- Deletion: WGY $\longrightarrow$ WY
- Insertion (dual): WY $\longrightarrow$ WGY
- Tandem duplication: WKY $\longrightarrow$ WKKY
- Tandem contraction (dual): WKKY $\longrightarrow$ WKY

Evolutionary model

- Substitution : WGY $\longrightarrow$ WKY
- Deletion: WGY $\longrightarrow$ WY
- Insertion (dual): WY $\longrightarrow$ WGY
- Tandem duplication: WKY $\longrightarrow$ WKKY
- Tandem contraction (dual): WKKY $\longrightarrow$ WKY

$\Rightarrow$ variation in their number of units

Example of an alignment of 2 maps



An alignment produced by *MS_Align* between maps *cccccccddddbaa* and *ccdddddbbcaccbbaa* with the costs $A = C = 1$, $I = D = 40$, $\mathcal{M}(a, b) = \mathcal{M}(a, d) = \mathcal{M}(b, d) = 20$, $\mathcal{M}(a, c) = \mathcal{M}(b, c) = \mathcal{M}(c, d) = 10$. Its cost is $14 \times A + \mathcal{M}(b, c) + \mathcal{M}(c, a) = 34$. Plain lines: matches, dashed lines: amplifications and contractions, bold lines: mutations.

Percentage of explained variance (VAF)

$$\text{VAF} = 1 - \frac{\sum_{(i,j):i<j}(D(i,j) - T(i,j))^2}{\sum_{(i,j):i<j}(D(i,j) - D_m)^2}$$

where

$D(i,j)$: alignment distance between i and j

$T(i,j)$: tree distance between i and j

D_m : average alignment distance over all pairs (i,j)

Value in $[0,1]$

Rate of elementary well designed quartets (Re)

For an internal edge e, for all quartets (i, j, k, l)
s.t. e splits (i, j) and (k, l):

$R(e) =$ percentage of these quartets satisfying

$$(D(i,j) + D(k,l)) < \min\left(D(i,l) - D(j,k)\right), \left(D(i,k) - D(j,l)\right)$$

where
$D(i, j)$: alignment distance between i and j

Value in $[0, 1]$

Results

Identical alleles

Locus	2	3	≥ 4	Total	$\neq$ origin	$\neq$ subspecies
MMS 24	7	4	1	27	8 (18)	0 (1)
MMS 26	6	2	3	36	10 (26)	5 (1)
MMS 30	9	2	1	38	8 (18)	2 (0)
MMS 80	10	8	0	44	10 (22)	0 (0)

Identical alleles

Locus	2	3	≥ 4	Total	$\neq$ origin	$\neq$ subspecies
MMS 24	7	4	1	27	8 (18)	0 (1)
MMS 26	6	2	3	36	10 (26)	5 (1)
MMS 30	9	2	1	38	8 (18)	2 (0)
MMS 80	10	8	0	44	10 (22)	0 (0)

Example

At MMS 30: DOM_BULG_Vlas_DBV, DOM_TUNI_Monastir_22MO, and SPR_MARO_Azzemour_9852
GYKKKGWGKoGGYWYKKoKKKYYYKG

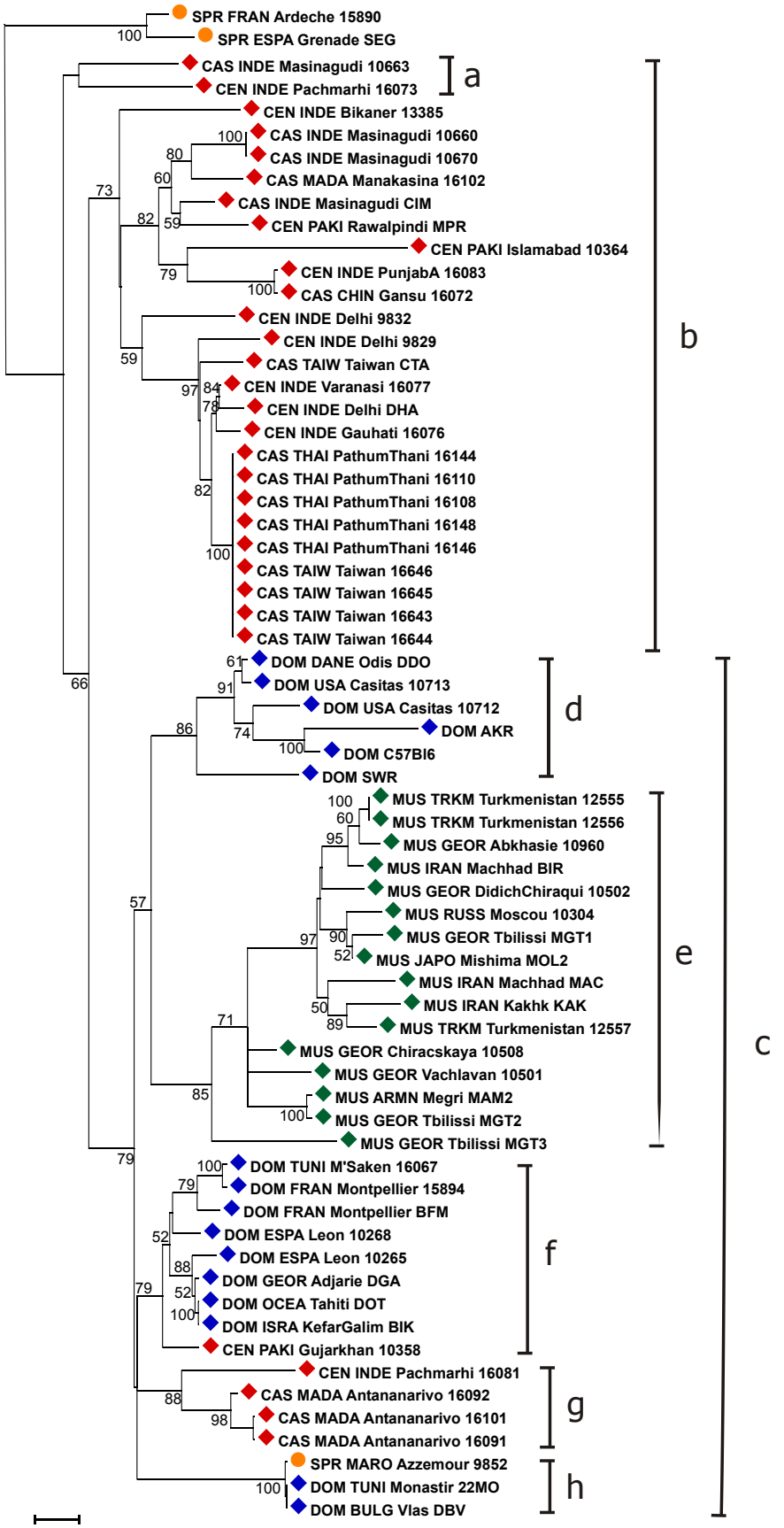
Identical alleles

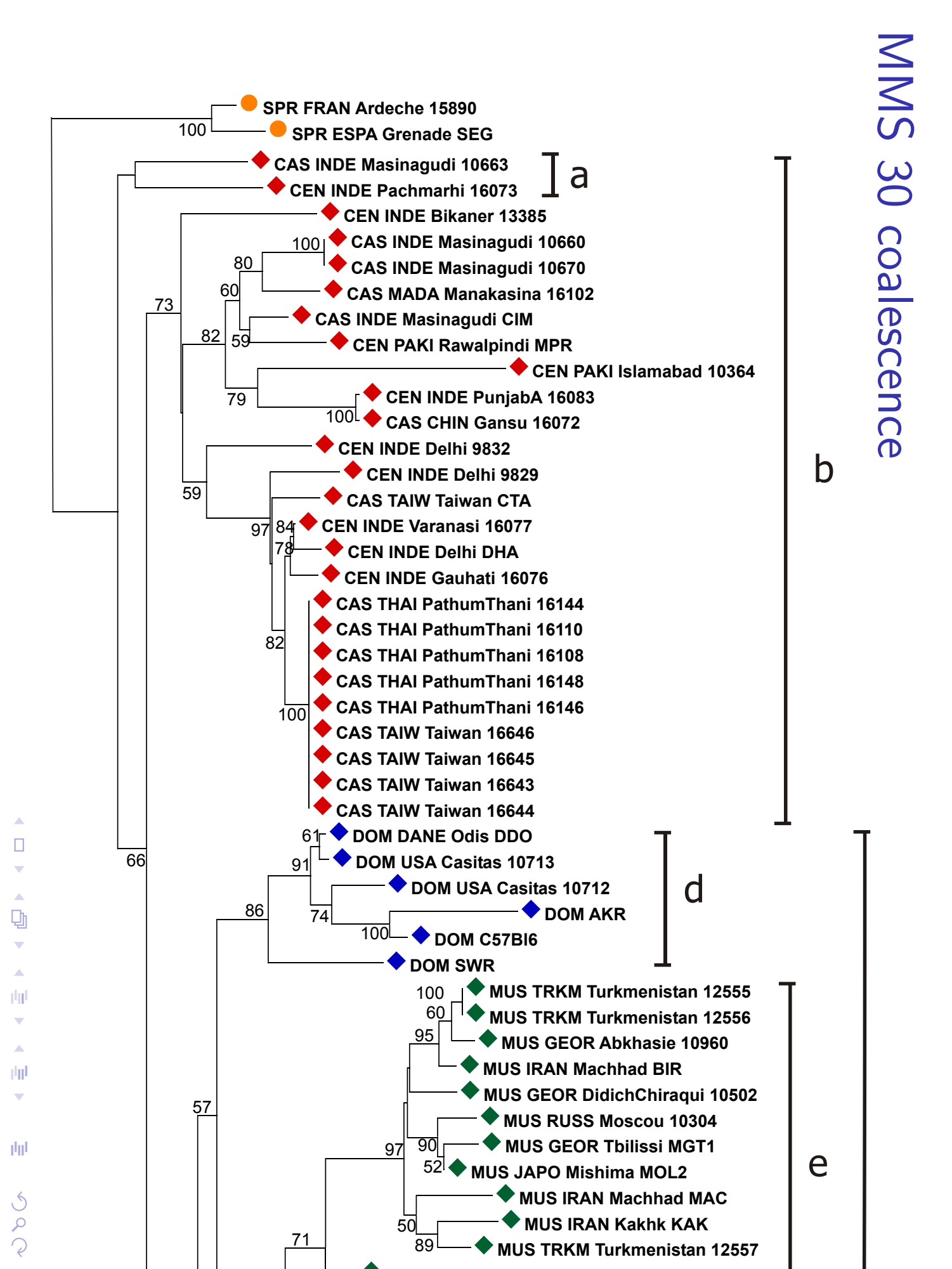
Locus	2	3	≥ 4	Total	$\neq$ origin	$\neq$ subspecies
MMS 24	7	4	1	27	8 (18)	0 (1)
MMS 26	6	2	3	36	10 (26)	5 (1)
MMS 30	9	2	1	38	8 (18)	2 (0)
MMS 80	10	8	0	44	10 (22)	0 (0)

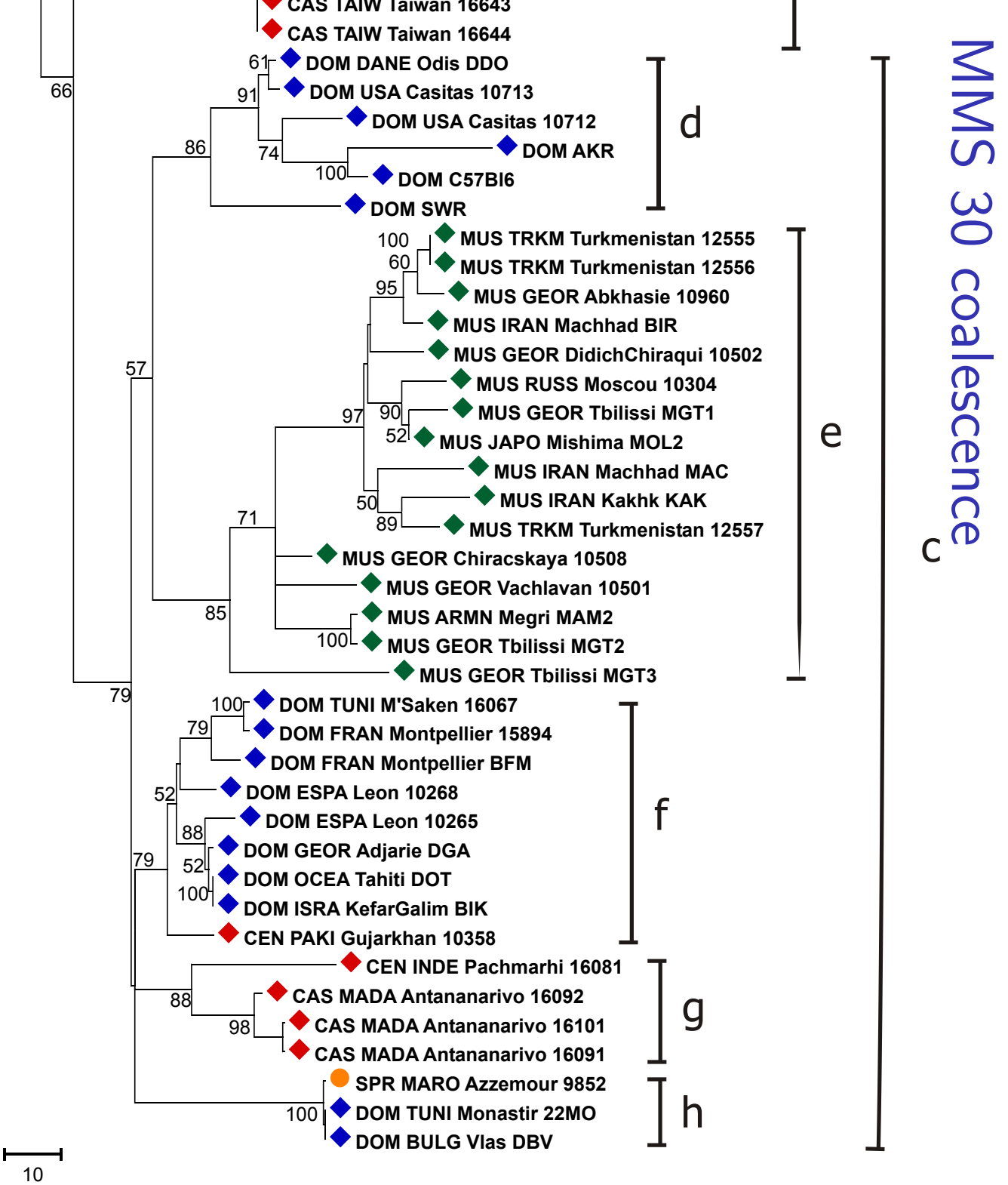
Example

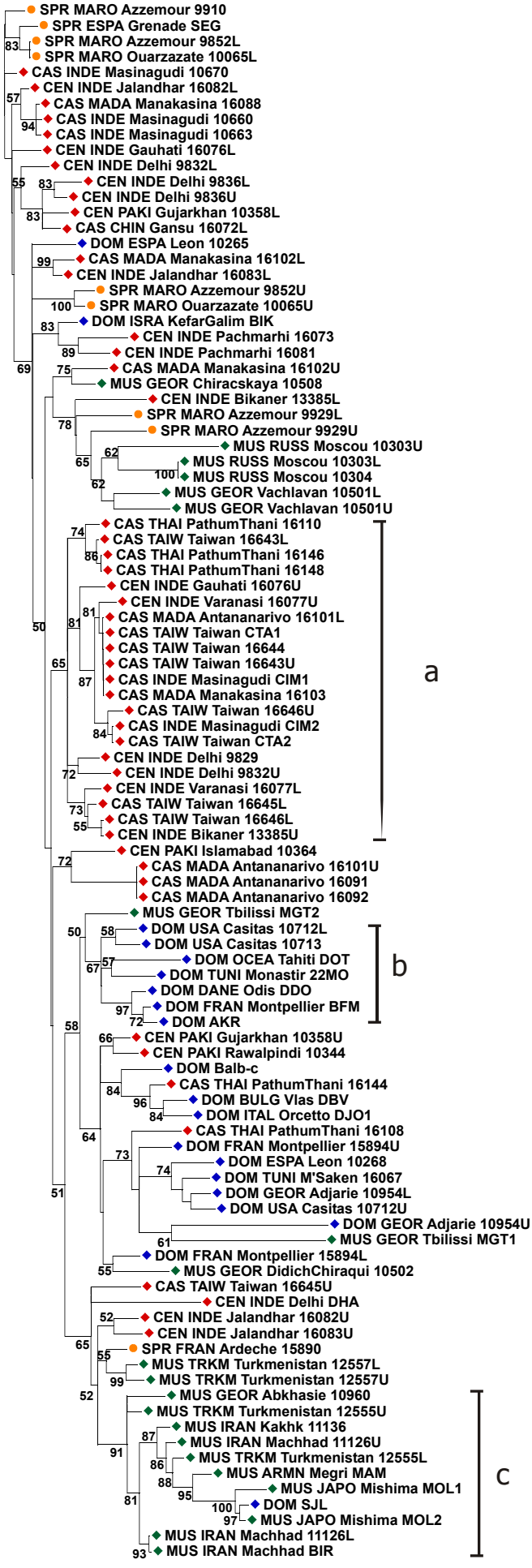
At MMS 30: DOM_BULG_Vlas_DBV, DOM_TUNI_Monastir_22MO, and SPR_MARO_Azzemour_9852
GYKKKGWGKoGGYWYKKoKKKYYYKG
At MMS 26: CEN_INDE_Dehli_DHA; DOM_OCEA_Tahiti_DOT
YGGGGGGGGAGGGAGAAGGYAAGGGGAAAAGAGAAGAAGGGG

MMS 30 coalescence

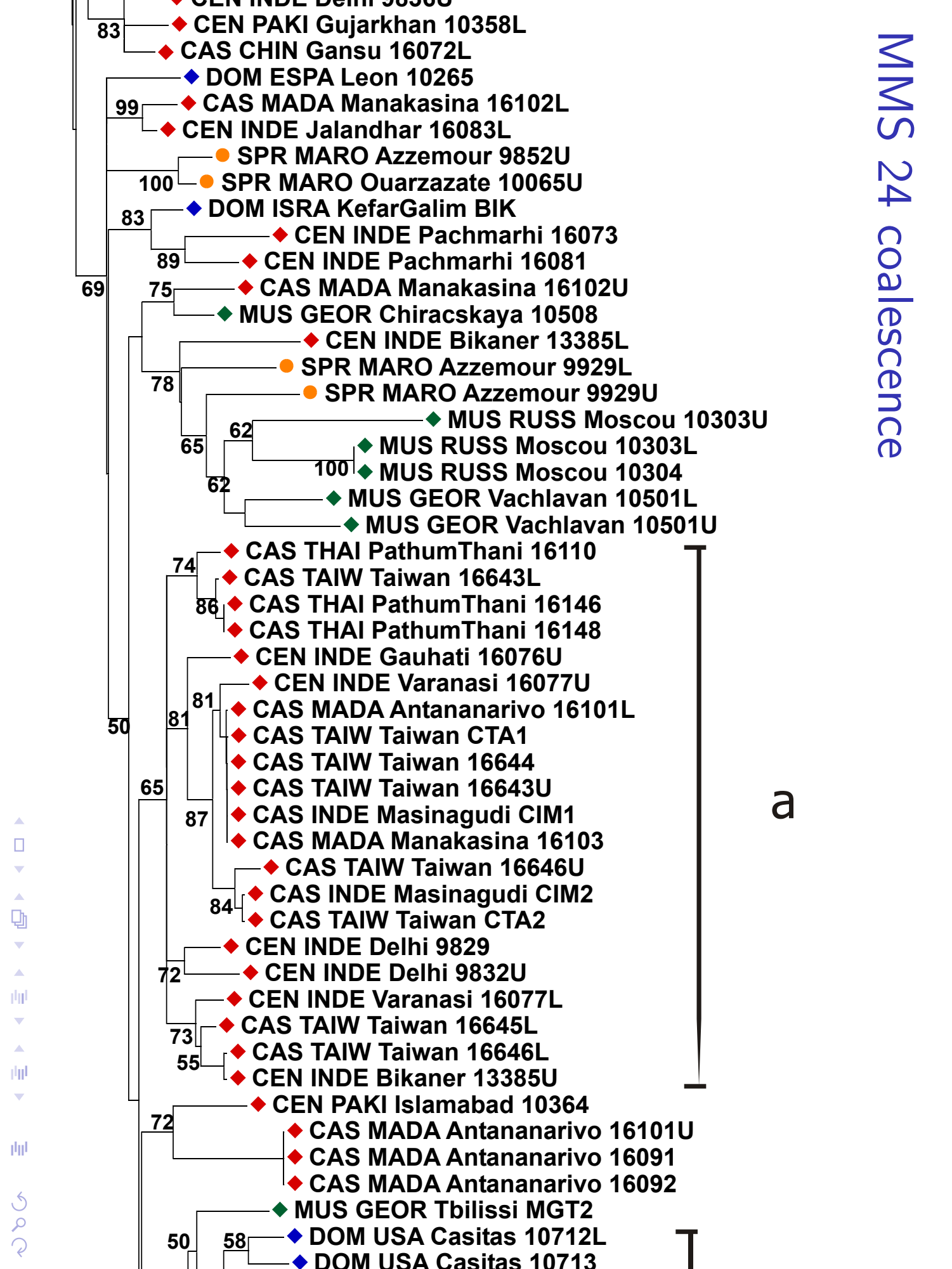


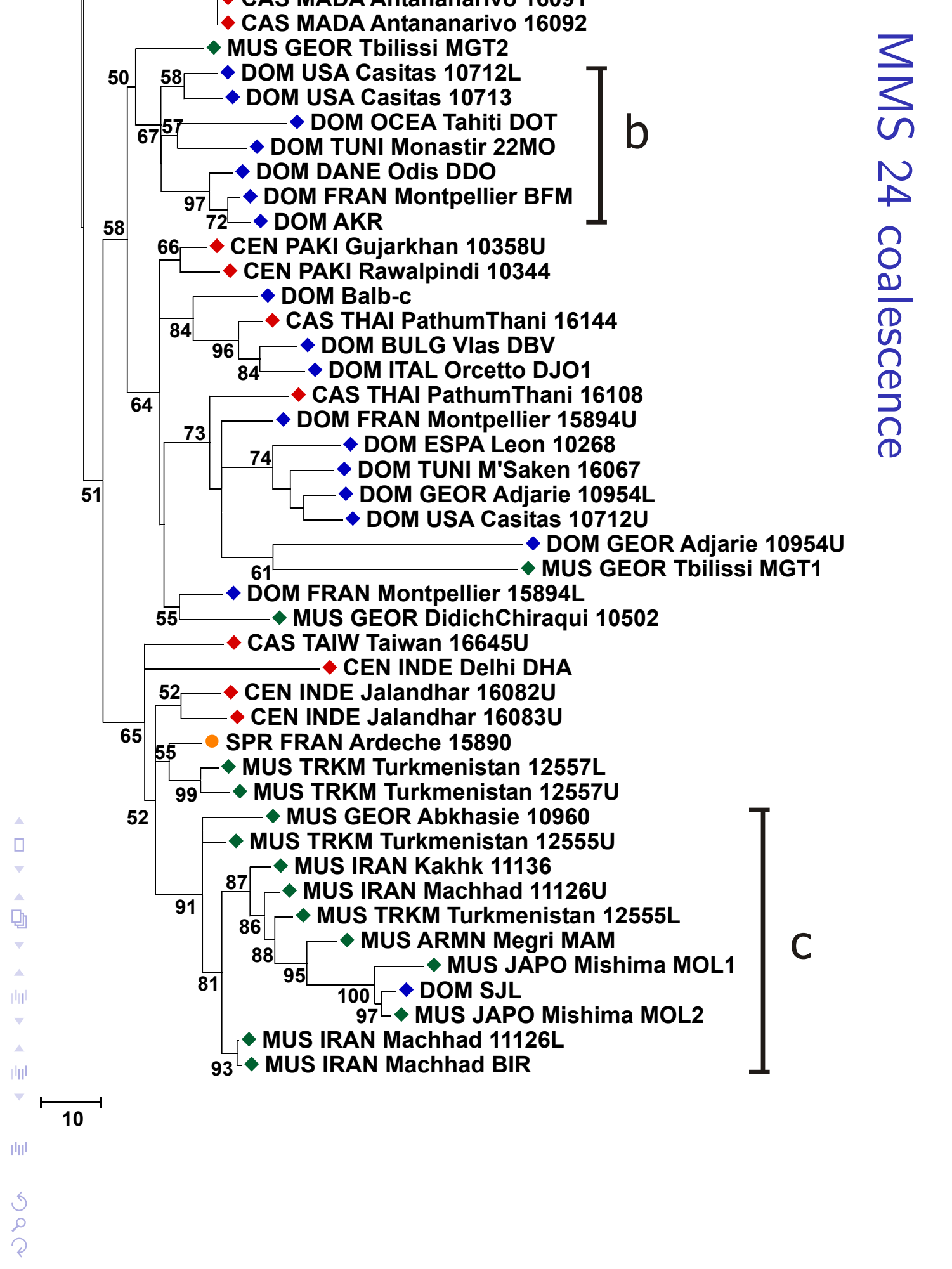

MMS 30 coalescence
SPR FRAN Ardeche 15890
SPR ESPA Grenade SEG
CAS INDE Masinagudi 10663
CEN INDE Pachmarhi 16073
CEN INDE Bikaner 13385
CAS INDE Masinagudi 10660
CAS INDE Masinagudi 10670
CAS MADA Manakasina 16102
CAS INDE Masinagudi CIM
CEN PAKI Rawalpindi MPR
CEN PAKI Islamabad 10364
CEN INDE PunjabA 16083
CAS CHIN Gansu 16072
CEN INDE Delhi 9832
CEN INDE Delhi 9829
CAS TAIW Taiwan CTA
CEN INDE Varanasi 16077
CEN INDE Delhi DHA
CEN INDE Gauhati 16076
CAS THAI PathumThani 16144
CAS THAI PathumThani 16110
CAS THAI PathumThani 16108
CAS THAI PathumThani 16148
CAS THAI PathumThani 16146
CAS TAIW Taiwan 16646
CAS TAIW Taiwan 16645
CAS TAIW Taiwan 16643
CAS TAIW Taiwan 16644
DOM DANE Odis DDO
DOM USA Casitas 10713
DOM USA Casitas 10712
DOM AKR
DOM C57Bl6
DOM SWR
MUS TRKM Turkmenistan 12555
MUS TRKM Turkmenistan 12556
MUS GEOR Abkhasie 10960
MUS IRAN Machhad BIR
MUS GEOR DidichChiraqui 10502
MUS RUSS Moscou 10304
MUS GEOR Tbilissi MGT1
MUS JAPO Mishima MOL2
MUS IRAN Machhad MAC
MUS IRAN Kakhk KAK
MUS TRKM Turkmenistan 12557
a
b
d
e

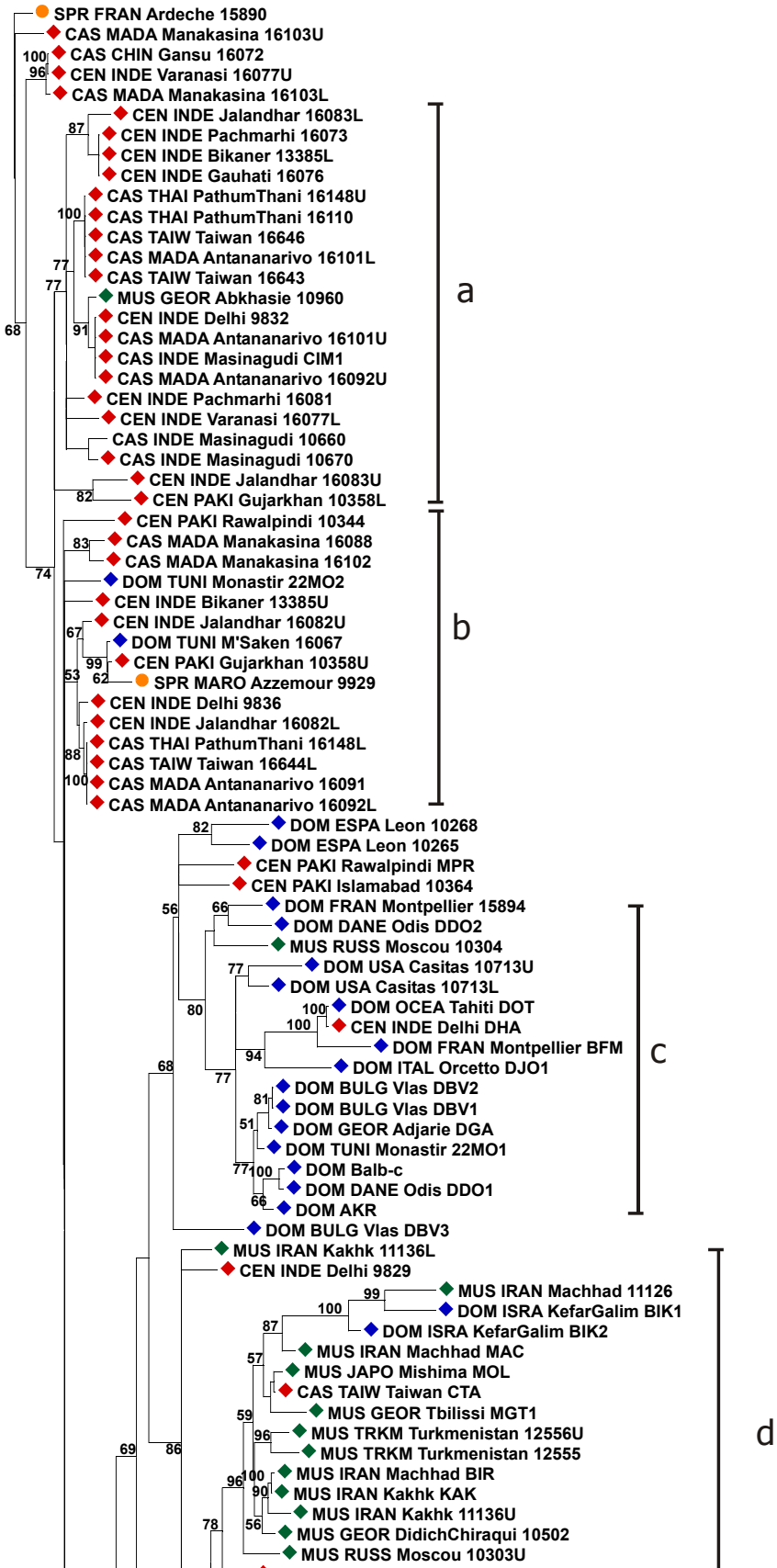

CAS TAIW Taiwan 16643
CAS TAIW Taiwan 16644
DOM DANE Odis DDO
DOM USA Casitas 10713
DOM USA Casitas 10712
DOM AKR
DOM C57Bl6
DOM SWR
MUS TRKM Turkmenistan 12555
MUS TRKM Turkmenistan 12556
MUS GEOR Abkhasie 10960
MUS IRAN Machhad BIR
MUS GEOR DidichChiraqui 10502
MUS RUSS Moscou 10304
MUS GEOR Tbilissi MGT1
MUS JAPO Mishima MOL2
MUS IRAN Machhad MAC
MUS IRAN Kakhk KAK
MUS TRKM Turkmenistan 12557
MUS GEOR Chiracskaya 10508
MUS GEOR Vachlavan 10501
MUS ARMN Megri MAM2
MUS GEOR Tbilissi MGT2
MUS GEOR Tbilissi MGT3
DOM TUNI M'Saken 16067
DOM FRAN Montpellier 15894
DOM FRAN Montpellier BFM
DOM ESPA Leon 10268
DOM ESPA Leon 10265
DOM GEOR Adjarie DGA
DOM OCEA Tahiti DOT
DOM ISRA KefarGalim BIK
CEN PAKI Gujarkhan 10358
CEN INDE Pachmarhi 16081
CAS MADA Antananarivo 16092
CAS MADA Antananarivo 16101
CAS MADA Antananarivo 16091
SPR MARO Azzemour 9852
DOM TUNI Monastir 22MO
DOM BULG Vlas DBV
d
e
f
g
h
c
MMS 30 coalescence
10

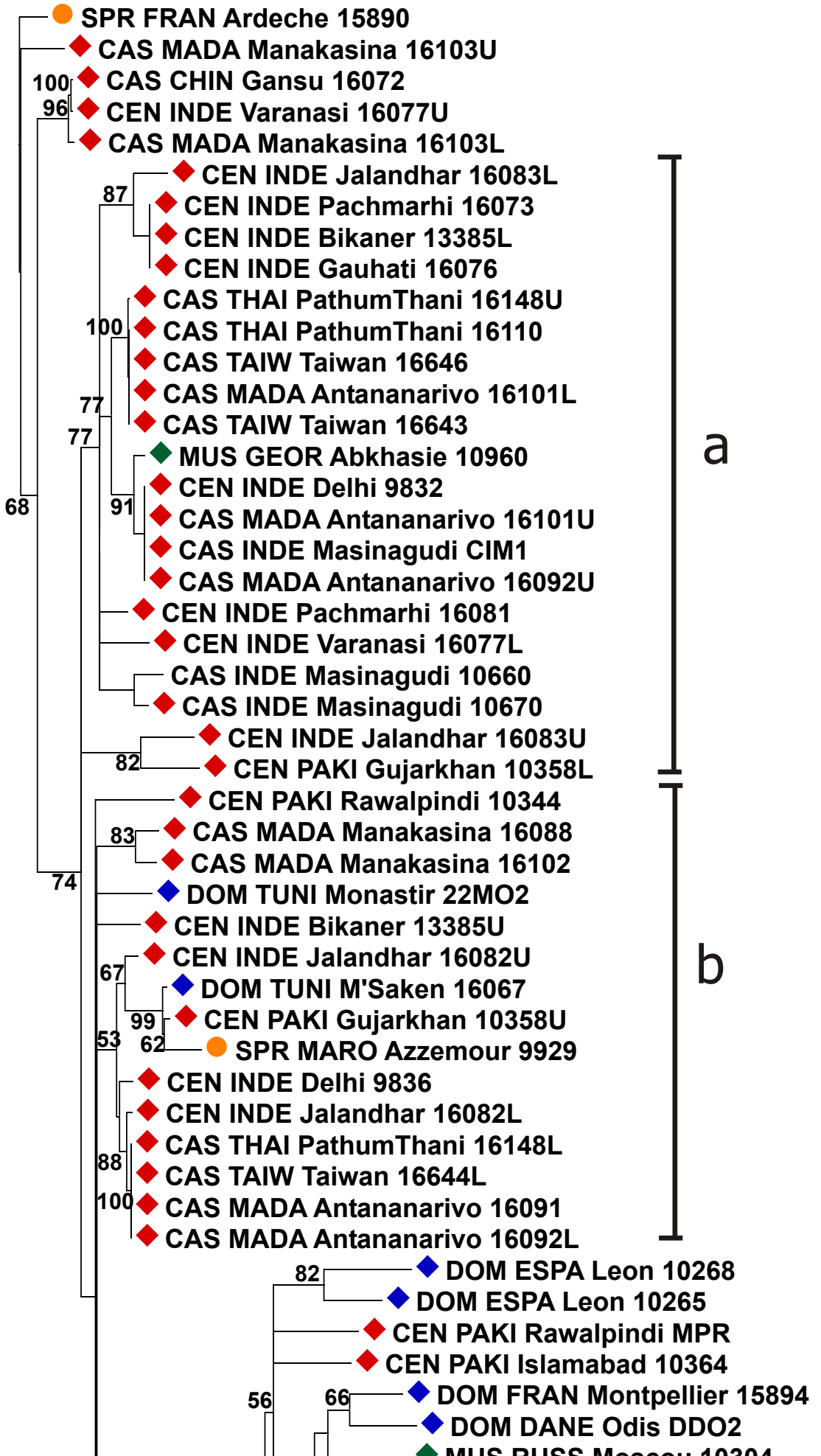


MMS 24 coalescence

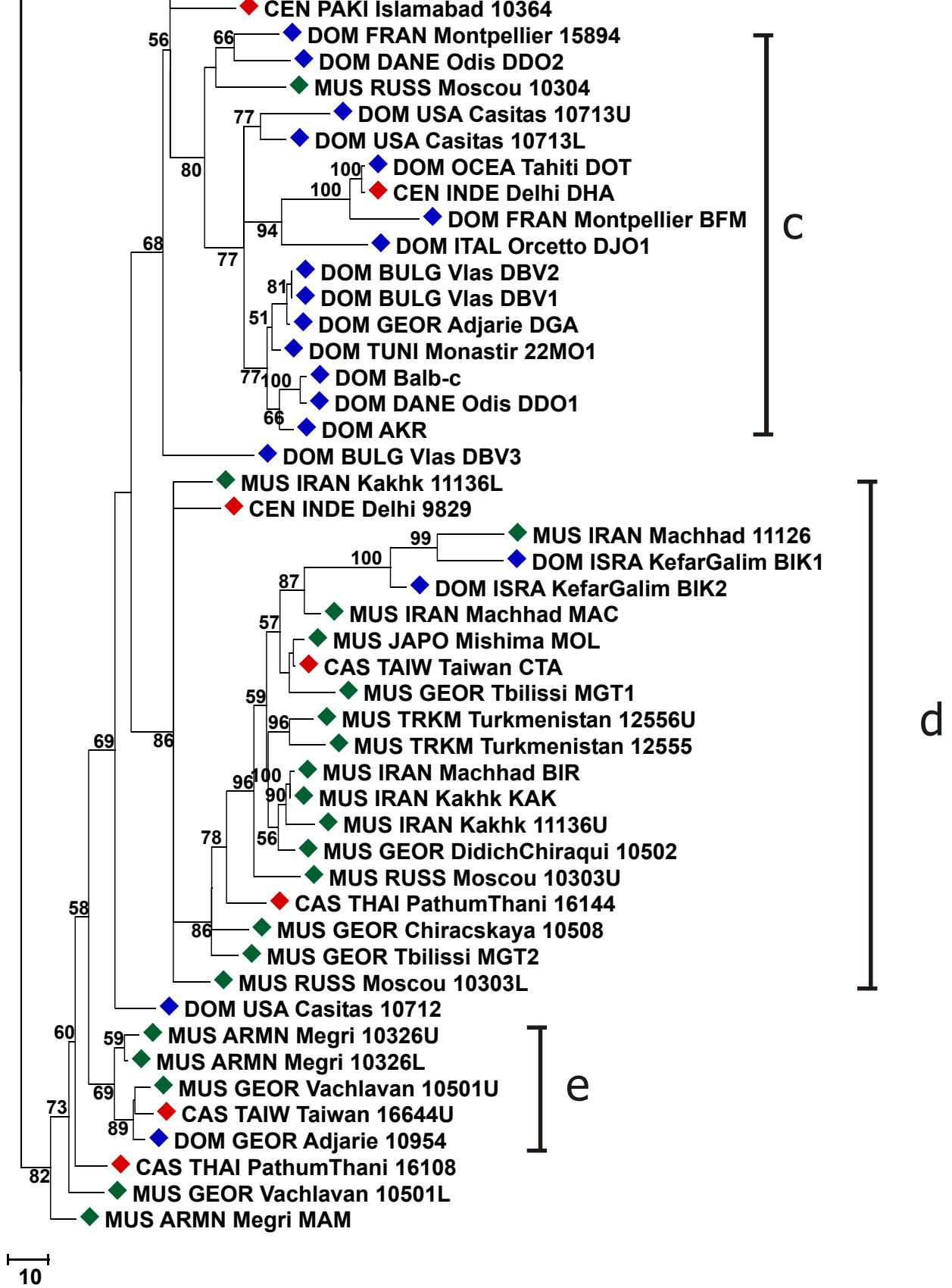

CEN PAKI Gujarkhan 10358L
CAS CHIN Gansu 16072L
DOM ESPA Leon 10265
CAS MADA Manakasina 16102L
CEN INDE Jalandhar 16083L
SPR MARO Azzemour 9852U
SPR MARO Ouarzazate 10065U
DOM ISRA KefarGalim BIK
CEN INDE Pachmarhi 16073
CEN INDE Pachmarhi 16081
CAS MADA Manakasina 16102U
MUS GEOR Chiracskaya 10508
CEN INDE Bikaner 13385L
SPR MARO Azzemour 9929L
SPR MARO Azzemour 9929U
MUS RUSS Moscou 10303U
MUS RUSS Moscou 10303L
MUS RUSS Moscou 10304
MUS GEOR Vachlavan 10501L
MUS GEOR Vachlavan 10501U
CAS THAI PathumThani 16110
CAS TAIW Taiwan 16643L
CAS THAI PathumThani 16146
CAS THAI PathumThani 16148
CEN INDE Gauhati 16076U
CEN INDE Varanasi 16077U
CAS MADA Antananarivo 16101L
CAS TAIW Taiwan CTA1
CAS TAIW Taiwan 16644
CAS TAIW Taiwan 16643U
CAS INDE Masinagudi CIM1
CAS MADA Manakasina 16103
CAS TAIW Taiwan 16646U
CAS INDE Masinagudi CIM2
CAS TAIW Taiwan CTA2
CEN INDE Delhi 9829
CEN INDE Delhi 9832U
CEN INDE Varanasi 16077L
CAS TAIW Taiwan 16645L
CAS TAIW Taiwan 16646L
CEN INDE Bikaner 13385U
CEN PAKI Islamabad 10364
CAS MADA Antananarivo 16101U
CAS MADA Antananarivo 16091
CAS MADA Antananarivo 16092
MUS GEOR Tbilissi MGT2
DOM USA Casitas 10712L
DOM USA Casitas 10713
a
MMS 24 coalescence


CAS MADA Antananarivo 16092
MUS GEOR Tbilissi MGT2
DOM USA Casitas 10712L
DOM USA Casitas 10713
DOM OCEA Tahiti DOT
DOM TUNI Monastir 22MO
DOM DANE Odis DDO
DOM FRAN Montpellier BFM
DOM AKR
b
CEN PAKI Gujarkhan 10358U
CEN PAKI Rawalpindi 10344
DOM Balb-c
CAS THAI PathumThani 16144
DOM BULG Vlas DBV
DOM ITAL Orcetto DJO1
CAS THAI PathumThani 16108
DOM FRAN Montpellier 15894U
DOM ESPA Leon 10268
DOM TUNI M'Saken 16067
DOM GEOR Adjarie 10954L
DOM USA Casitas 10712U
DOM GEOR Adjarie 10954U
MUS GEOR Tbilissi MGT1
DOM FRAN Montpellier 15894L
MUS GEOR DidichChiraqui 10502
CAS TAIW Taiwan 16645U
CEN INDE Delhi DHA
CEN INDE Jalandhar 16082U
CEN INDE Jalandhar 16083U
SPR FRAN Ardeche 15890
MUS TRKM Turkmenistan 12557L
MUS TRKM Turkmenistan 12557U
MUS GEOR Abkhasie 10960
MUS TRKM Turkmenistan 12555U
MUS IRAN Kakhk 11136
MUS IRAN Machhad 11126U
MUS TRKM Turkmenistan 12555L
MUS ARMN Megri MAM
MUS JAPO Mishima MOL1
DOM SJL
MUS JAPO Mishima MOL2
MUS IRAN Machhad 11126L
MUS IRAN Machhad BIR
c
10
MMS 24 coalescence

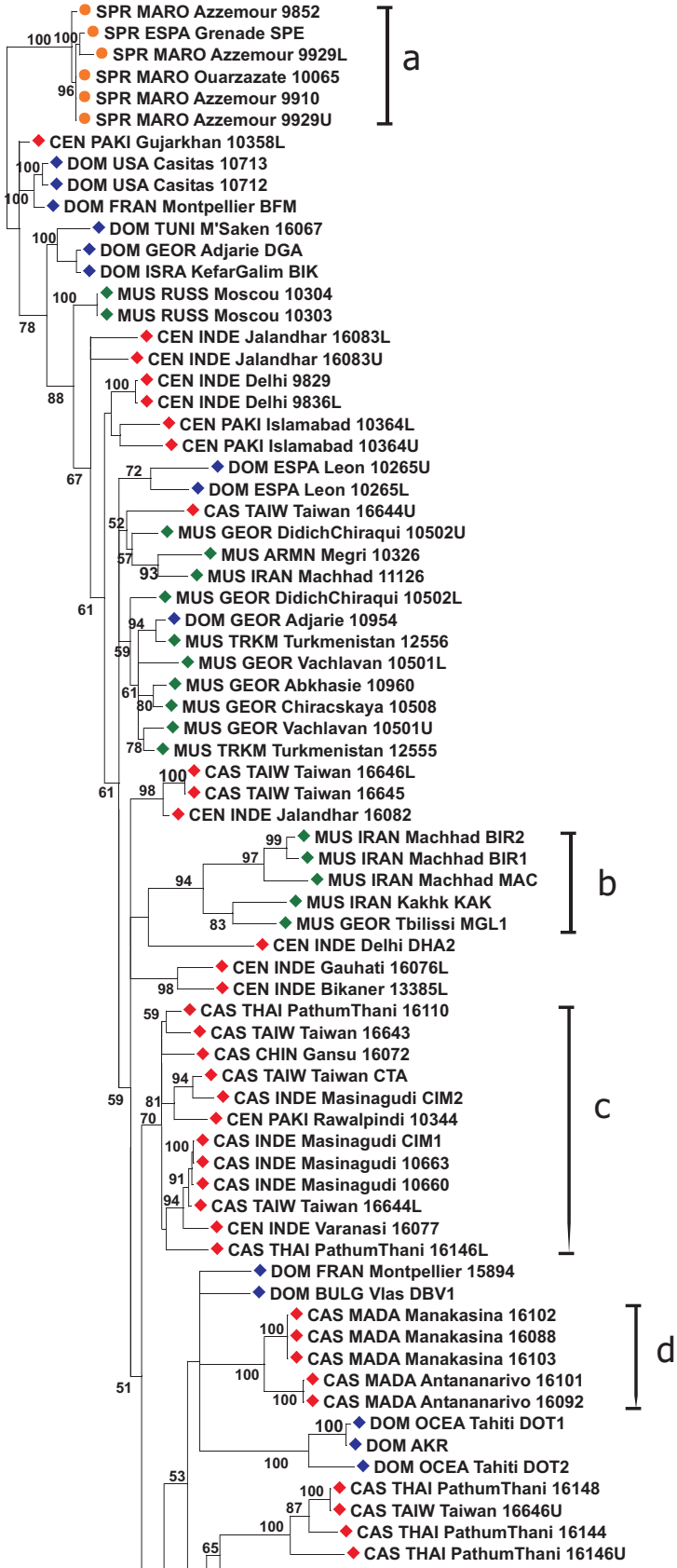

SPR FRAN Ardeche 15890
CAS MADA Manakasina 16103U
CAS CHIN Gansu 16072
CEN INDE Varanasi 16077U
CAS MADA Manakasina 16103L
CEN INDE Jalandhar 16083L
CEN INDE Pachmarhi 16073
CEN INDE Bikaner 13385L
CEN INDE Gauhati 16076
CAS THAI PathumThani 16148U
CAS THAI PathumThani 16110
CAS TAIW Taiwan 16646
CAS MADA Antananarivo 16101L
CAS TAIW Taiwan 16643
MUS GEOR Abkhasie 10960
CEN INDE Delhi 9832
CAS MADA Antananarivo 16101U
CAS INDE Masinagudi CIM1
CAS MADA Antananarivo 16092U
CEN INDE Pachmarhi 16081
CEN INDE Varanasi 16077L
CAS INDE Masinagudi 10660
CAS INDE Masinagudi 10670
CEN INDE Jalandhar 16083U
CEN PAKI Gujarkhan 10358L
a
CEN PAKI Rawalpindi 10344
CAS MADA Manakasina 16088
CAS MADA Manakasina 16102
DOM TUNI Monastir 22MO2
CEN INDE Bikaner 13385U
CEN INDE Jalandhar 16082U
DOM TUNI M'Saken 16067
CEN PAKI Gujarkhan 10358U
SPR MARO Azzemour 9929
CEN INDE Delhi 9836
CEN INDE Jalandhar 16082L
CAS THAI PathumThani 16148L
CAS TAIW Taiwan 16644L
CAS MADA Antananarivo 16091
CAS MADA Antananarivo 16092L
b
DOM ESPA Leon 10268
DOM ESPA Leon 10265
CEN PAKI Rawalpindi MPR
CEN PAKI Islamabad 10364
DOM FRAN Montpellier 15894
DOM DANE Odis DDO2
MUS RUSS Moscou 10304
DOM USA Casitas 10713U
DOM USA Casitas 10713L
DOM OCEA Tahiti DOT
CEN INDE Delhi DHA
DOM FRAN Montpellier BFM
DOM ITAL Orcetto DJO1
DOM BULG Vlas DBV2
DOM BULG Vlas DBV1
DOM GEOR Adjarie DGA
DOM TUNI Monastir 22MO1
DOM Balb-c
DOM DANE Odis DDO1
DOM AKR
c
DOM BULG Vlas DBV3
MUS IRAN Kakhk 11136L
CEN INDE Delhi 9829
MUS IRAN Machhad 11126
DOM ISRA KefarGalim BIK1
DOM ISRA KefarGalim BIK2
MUS IRAN Machhad MAC
MUS JAPO Mishima MOL
CAS TAIW Taiwan CTA
MUS GEOR Tbilissi MGT1
MUS TRKM Turkmenistan 12556U
MUS TRKM Turkmenistan 12555
MUS IRAN Machhad BIR
MUS IRAN Kakhk KAK
MUS IRAN Kakhk 11136U
MUS GEOR DidichChiraqui 10502
MUS RUSS Moscou 10303U
d

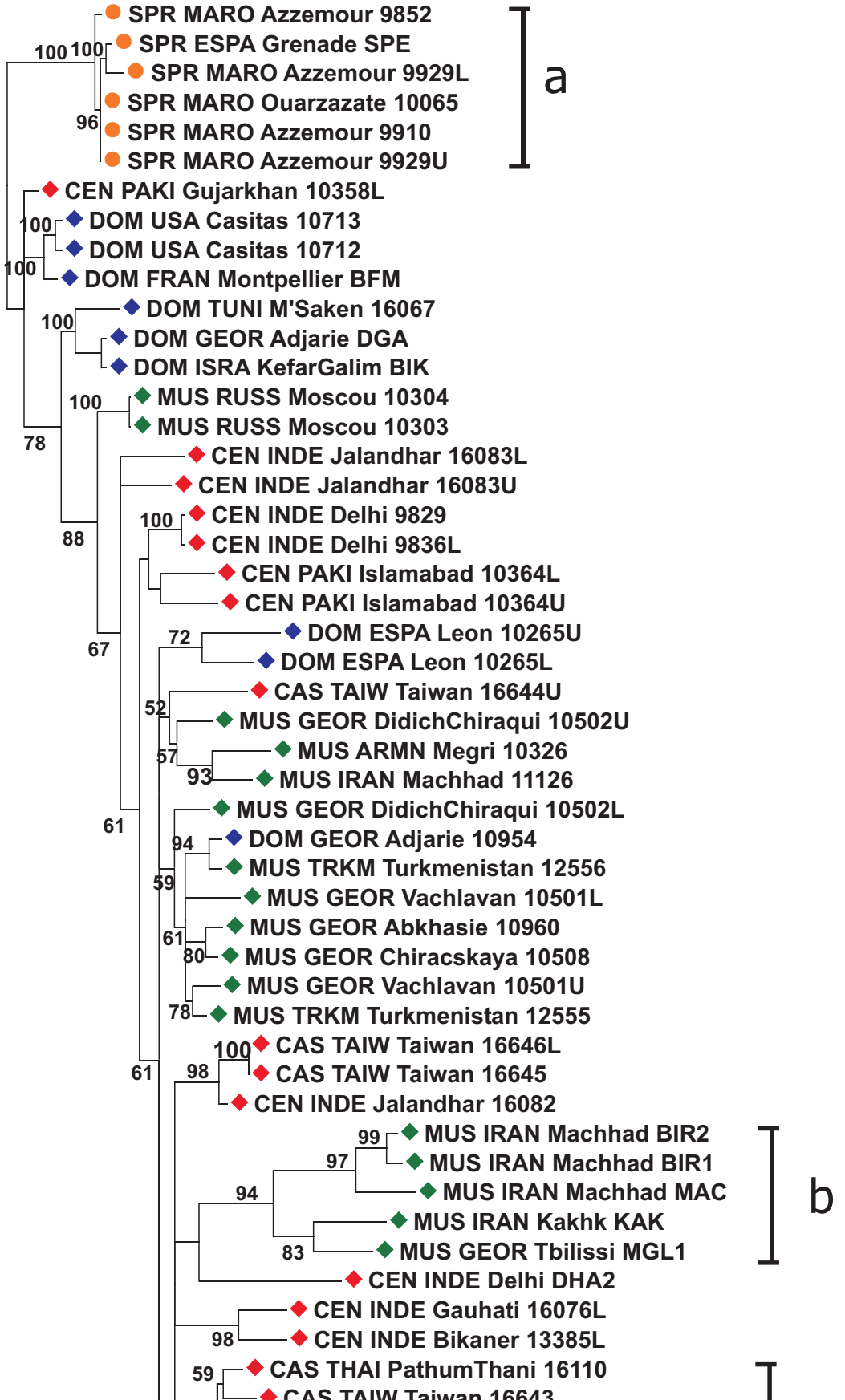


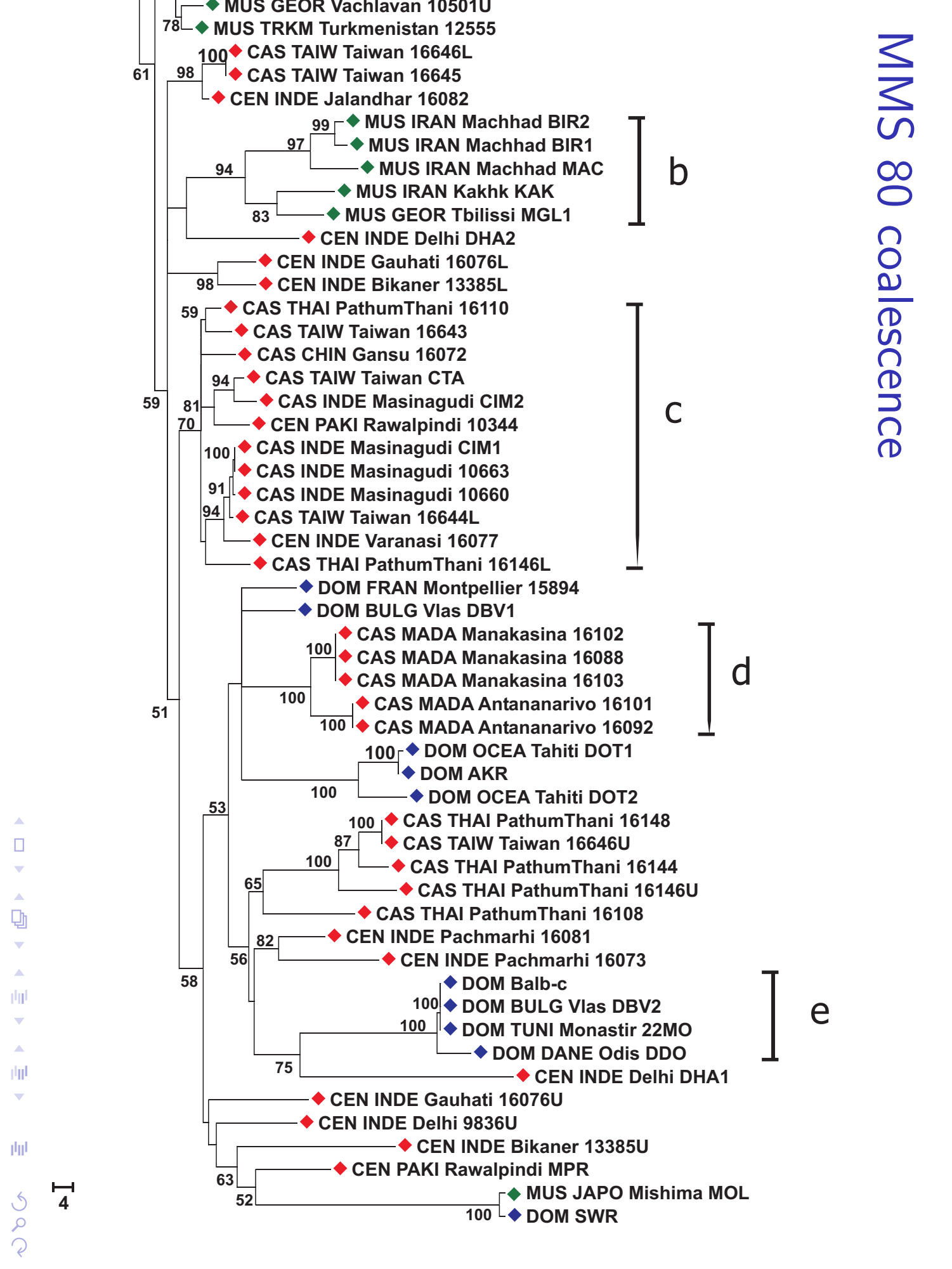
MMS 26 coalescence



MMS 26 coalescence


SPR MARO Azzemour 9852
SPR ESPA Grenade SPE
SPR MARO Azzemour 9929L
SPR MARO Ouarzazate 10065
SPR MARO Azzemour 9910
SPR MARO Azzemour 9929U
a
CEN PAKI Gujarkhan 10358L
DOM USA Casitas 10713
DOM USA Casitas 10712
DOM FRAN Montpellier BFM
DOM TUNI M'Saken 16067
DOM GEOR Adjarie DGA
DOM ISRA KefarGalim BIK
MUS RUSS Moscou 10304
MUS RUSS Moscou 10303
CEN INDE Jalandhar 16083L
CEN INDE Jalandhar 16083U
CEN INDE Delhi 9829
CEN INDE Delhi 9836L
CEN PAKI Islamabad 10364L
CEN PAKI Islamabad 10364U
DOM ESPA Leon 10265U
DOM ESPA Leon 10265L
CAS TAIW Taiwan 16644U
MUS GEOR DidichChiraqui 10502U
MUS ARMN Megri 10326
MUS IRAN Machhad 11126
MUS GEOR DidichChiraqui 10502L
DOM GEOR Adjarie 10954
MUS TRKM Turkmenistan 12556
MUS GEOR Vachlavan 10501L
MUS GEOR Abkhasie 10960
MUS GEOR Chiracskaya 10508
MUS GEOR Vachlavan 10501U
MUS TRKM Turkmenistan 12555
CAS TAIW Taiwan 16646L
CAS TAIW Taiwan 16645
CEN INDE Jalandhar 16082
MUS IRAN Machhad BIR2
MUS IRAN Machhad BIR1
MUS IRAN Machhad MAC
MUS IRAN Kakhk KAK
MUS GEOR Tbilissi MGL1
b
CEN INDE Delhi DHA2
CEN INDE Gauhati 16076L
CEN INDE Bikaner 13385L
CAS THAI PathumThani 16110
CAS TAIW Taiwan 16643
CAS CHIN Gansu 16072
CAS TAIW Taiwan CTA
CAS INDE Masinagudi CIM2
CEN PAKI Rawalpindi 10344
CAS INDE Masinagudi CIM1
CAS INDE Masinagudi 10663
CAS INDE Masinagudi 10660
CAS TAIW Taiwan 16644L
CEN INDE Varanasi 16077
CAS THAI PathumThani 16146L
c
DOM FRAN Montpellier 15894
DOM BULG Vlas DBV1
CAS MADA Manakasina 16102
CAS MADA Manakasina 16088
CAS MADA Manakasina 16103
CAS MADA Antananarivo 16101
CAS MADA Antananarivo 16092
d
DOM OCEA Tahiti DOT1
DOM AKR
DOM OCEA Tahiti DOT2
CAS THAI PathumThani 16148
CAS TAIW Taiwan 16646U
CAS THAI PathumThani 16144
CAS THAI PathumThani 16146U


SPR MARO Azzemour 9852
SPR ESPA Grenade SPE
SPR MARO Azzemour 9929L
SPR MARO Ouarzazate 10065
SPR MARO Azzemour 9910
SPR MARO Azzemour 9929U
a
CEN PAKI Gujarkhan 10358L
DOM USA Casitas 10713
DOM USA Casitas 10712
DOM FRAN Montpellier BFM
DOM TUNI M'Saken 16067
DOM GEOR Adjarie DGA
DOM ISRA KefarGalim BIK
MUS RUSS Moscou 10304
MUS RUSS Moscou 10303
CEN INDE Jalandhar 16083L
CEN INDE Jalandhar 16083U
CEN INDE Delhi 9829
CEN INDE Delhi 9836L
CEN PAKI Islamabad 10364L
CEN PAKI Islamabad 10364U
DOM ESPA Leon 10265U
DOM ESPA Leon 10265L
CAS TAIW Taiwan 16644U
MUS GEOR DidichChiraqui 10502U
MUS ARMN Megri 10326
MUS IRAN Machhad 11126
MUS GEOR DidichChiraqui 10502L
DOM GEOR Adjarie 10954
MUS TRKM Turkmenistan 12556
MUS GEOR Vachlavan 10501L
MUS GEOR Abkhasie 10960
MUS GEOR Chiracskaya 10508
MUS GEOR Vachlavan 10501U
MUS TRKM Turkmenistan 12555
CAS TAIW Taiwan 16646L
CAS TAIW Taiwan 16645
CEN INDE Jalandhar 16082
MUS IRAN Machhad BIR2
MUS IRAN Machhad BIR1
MUS IRAN Machhad MAC
MUS IRAN Kakhk KAK
MUS GEOR Tbilissi MGL1
b
CEN INDE Delhi DHA2
CEN INDE Gauhati 16076L
CEN INDE Bikaner 13385L
CAS THAI PathumThani 16110
CAS TAIW Taiwan 16643


MMS 80 coalescence
MUS GEOR Vachlavan 10501U
MUS TRKM Turkmenistan 12555
CAS TAIW Taiwan 16646L
CAS TAIW Taiwan 16645
CEN INDE Jalandhar 16082
MUS IRAN Machhad BIR2
MUS IRAN Machhad BIR1
MUS IRAN Machhad MAC
MUS IRAN Kakhk KAK
MUS GEOR Tbilissi MGL1
CEN INDE Delhi DHA2
CEN INDE Gauhati 16076L
CEN INDE Bikaner 13385L
CAS THAI PathumThani 16110
CAS TAIW Taiwan 16643
CAS CHIN Gansu 16072
CAS TAIW Taiwan CTA
CAS INDE Masinagudi CIM2
CEN PAKI Rawalpindi 10344
CAS INDE Masinagudi CIM1
CAS INDE Masinagudi 10663
CAS INDE Masinagudi 10660
CAS TAIW Taiwan 16644L
CEN INDE Varanasi 16077
CAS THAI PathumThani 16146L
DOM FRAN Montpellier 15894
DOM BULG Vlas DBV1
CAS MADA Manakasina 16102
CAS MADA Manakasina 16088
CAS MADA Manakasina 16103
CAS MADA Antananarivo 16101
CAS MADA Antananarivo 16092
DOM OCEA Tahiti DOT1
DOM AKR
DOM OCEA Tahiti DOT2
CAS THAI PathumThani 16148
CAS TAIW Taiwan 16646U
CAS THAI PathumThani 16144
CAS THAI PathumThani 16146U
CAS THAI PathumThani 16108
CEN INDE Pachmarhi 16081
CEN INDE Pachmarhi 16073
DOM Balb-c
DOM BULG Vlas DBV2
DOM TUNI Monastir 22MO
DOM DANE Odis DDO
CEN INDE Delhi DHA1
CEN INDE Gauhati 16076U
CEN INDE Delhi 9836U
CEN INDE Bikaner 13385U
CEN PAKI Rawalpindi MPR
MUS JAPO Mishima MOL
DOM SWR
b
c
d
e
4

Intruders

Locus	ID	clade	Average distance to CAS	CEN	DOM	MUS
24	CAS THAI Pathumtani 16108	DOM	52	53	40	61
24	CAS THAI Pathumtani 16144	DOM	45	47	37	60
24	DOM SJL	MUS	71	74	74	53
26	MUS GEOR Abkhasie 10960	CAS	23	34	84	72
26	CEN INDE Dehli DHA	DOM	99	96	58	91
26	MUS RUSS Moscou 10304	DOM	69	62	60	92
26	CAS TAIW Taiwan CTA	MUS	77	80	73	42
80	DOM GEOR Adjarie 10954	MUS	37	39	56	28
80	DOM BULG Vlas DBV1	CAS	40	50	50	51
80	DOM FRAN Montpellier 15894	CAS	42	49	51	50

MMS24 intruders alignments

```
DOM_ESPA_Leon     CTCCC~CCT~~oTCTTCoTC~oTCToT~~oTTCCCC~~
DOM_USA_Casit     CTCCo~CCo~oo~CTTCoTC~oTCTo~~ooTTCCCC~~
DOM_GEOR_Adja     CTCCT~CCTT~o~CTTCoTCCoTCTo~~ooTTCCCC~~
DOM_TUNI_M'Sa     CTCCT~CCT~~oTCTTCoTC~oTCTo~~ooTTC~~~~~
DOM_FRAN_Mont     CTCCT~CCT~~oTCTTC~~~~oTCTT~~~oTTCCCC~~
CAS_THAI_Path     CTCCT~CCT~~oTCTTC~~~~oTCoT~~~CT~~~~~~~
CAS_THAI_Path     CTCCT~CCTT~~TCTTCTTC~TTCTTTT~~TTC~~~~~
DOM_BULG_Vlas     CTCCTTC~TT~~TCTTCTTC~TTCTTTT~~TTCCCCTC
DOM_ITAL_Orce     CTCCT~CCTT~~TCTTCTTC~TTCTTT~~oTTCCCCT~
                  ****  *..    ****     **.     *..
```

MMS30 intruders alignments

```
DOM ESPA   GYKKKWGKLKYLWKGWGKoGGYWYKKoKKK~YYY~KG
DOM TURQ   GYKKKWGKWKYoWKGWGKoGGYLYKKoKKK~YYY~KG
DOM TURQ   GYKKKWGKWKYoWKGWGKoGGYLYKKoK~~~YYY~KG
DOM GEOR   GYKKKWGKLKYoWKGWGKoGGYWYKKoK~~~YYY~KG
DOM ISRA   GYKKKWGKLKYoWKGWGKoGGYWYKKoKKK~YYY~KG
DOM ESPA   GYKKKWGKLKWoWKGWGKoGGYWYKKoKKK~YYY~KG
DOM FRAN   GYKKKWGKLKYLWKGWGKWGGYWYKKoKKK~YY~~~G
DOM FRAN   GYKKKWGKLKYoWKGWGKWGGYYYKKoKKK~YY~~~G
DOM ITAL   GYKKKWGKLKYoWKGWGKoGGYWYKKoKKK~YY~~~G
DOM TURQ   GYKKKWGKLKYoWKGWGLWGGYWYKKoKKK~YYY~KG
CEN PAKI   GYKKKWGKoKYKWKGWGKoGGYWYKKoKKK~YYY~KG
CEN INDE   GYKKKWGLWKYooKoWGLWGGYWYKKoKKKKYY~~KG
CAS MADA   GYKKK-GLLKYLoKGWGLoGGKWKKKoKKKKYY~~KG
CAS MADA   GYKKKWGLLKYLoKGWGLoGGKWKKKoKKKKYY~~KG
           *****...  ..  .*.**  .**.     **    **   .*
```

MMS30 intruders alignments

```
DOM ESPA   GYKKKWGKLKYLWKGWGKoGGYWYKKoKKK~YYY~KG
DOM TURQ   GYKKKWGKWKYoWKGWGKoGGYLYKKoKKK~YYY~KG
DOM TURQ   GYKKKWGKWKYoWKGWGKoGGYLYKKoK~~~YYY~KG
DOM GEOR   GYKKKWGKLKYoWKGWGKoGGYWYKKoK~~~YYY~KG
DOM ISRA   GYKKKWGKLKYoWKGWGKoGGYWYKKoKKK~YYY~KG
DOM ESPA   GYKKKWGKLKWoWKGWGKoGGYWYKKoKKK~YYY~KG
DOM FRAN   GYKKKWGKLKYLWKGWGKWGGYWYKKoKKK~YY~~~G
DOM FRAN   GYKKKWGKLKYoWKGWGKWGGYYYKKoKKK~YY~~~G
DOM ITAL   GYKKKWGKLKYoWKGWGKoGGYWYKKoKKK~YY~~~G
DOM TURQ   GYKKKWGKLKYoWKGWGLWGGYWYKKoKKK~YYY~KG
CEN PAKI   GYKKKWGKoKYKWKGWGKoGGYWYKKoKKK~YYY~KG
CEN INDE   GYKKKWGLWKYooKoWGLWGGYWYKKoKKKKYY~~KG
CAS MADA   GYKKK-GLLKYLoKGWGLoGGKWKKKoKKKKYY~~KG
CAS MADA   GYKKKWGLLKYLoKGWGLoGGKWKKKoKKKKYY~~KG
           *****...  ..  .*.**  .**.     **    **   .*
```

MMS30 intruders alignments

```
DOM ESPA   GYKKKWGKLKYLWKGWGKoGGYWYKKoKKK~YYY~KG
DOM TURQ   GYKKKWGKWKYoWKGWGKoGGYLYKKoKKK~YYY~KG
DOM TURQ   GYKKKWGKWKYoWKGWGKoGGYLYKKoK~~~YYY~KG
DOM GEOR   GYKKKWGKLKYoWKGWGKoGGYWYKKoK~~~YYY~KG
DOM ISRA   GYKKKWGKLKYoWKGWGKoGGYWYKKoKKK~YYY~KG
DOM ESPA   GYKKKWGKLKWoWKGWGKoGGYWYKKoKKK~YYY~KG
DOM FRAN   GYKKKWGKLKYLWKGWGKWGGYWYKKoKKK~YY~~~G
DOM FRAN   GYKKKWGKLKYoWKGWGKWGGYYYKKoKKK~YY~~~G
DOM ITAL   GYKKKWGKLKYoWKGWGKoGGYWYKKoKKK~YY~~~G
DOM TURQ   GYKKKWGKLKYoWKGWGLWGGYWYKKoKKK~YYY~KG
CEN PAKI   GYKKKWGKoKYKWKGWGKoGGYWYKKoKKK~YYY~KG
CEN INDE   GYKKKWGLWKYooKoWGLWGGYWYKKoKKKKYY~~KG
CAS MADA   GYKKK-GLLKYLoKGWGLoGGKWKKKoKKKKYY~~KG
CAS MADA   GYKKKWGLLKYLoKGWGLoGGKWKKKoKKKKYY~~KG
           *****... .. .*.** .**.    **   **   .*
```

Conclusion

Conclusion

- MS analysis revealed:
  1. species-wide genetic flow
  2. past and present exchanges within the species range

Conclusion

- MS analysis revealed:
  1. species-wide genetic flow
  2. past and present exchanges within the species range
- when handled with appropriate comparison algorithms [Bérard, Rivals, 2003]

Conclusion

- MS analysis revealed:
  1. species-wide genetic flow
  2. past and present exchanges within the species range
- when handled with appropriate comparison algorithms [Bérard, Rivals, 2003]
- MS versus SNPs

Conclusion

- MS analysis revealed:
  1. species-wide genetic flow
  2. past and present exchanges within the species range
- when handled with appropriate comparison algorithms [Bérard, Rivals, 2003]
- MS versus SNPs and micro-satellites

Conclusion

- MS analysis revealed:
  1. species-wide genetic flow
  2. past and present exchanges within the species range
- when handled with appropriate comparison algorithms [Bérard, Rivals, 2003]
- MS versus SNPs and micro-satellites
- Mice genomes: set of interrelated gene pools, still able to exchange genes, even on chromosome X
- Difficult to find genes important for speciation

Publications and algorithms

- S. Bérard, E. Rivals, *Comparison of Minisatellites*, *J. of Computational Biology*, p. 357-372, vol. 10(3-4), 2003.

- S. Bérard, F. Nicolas, J. Buard, O. Gascuel, E. Rivals, *A Fast and Specific Alignment Method for Minisatellite Maps*, *Evolutionary Bioinformatics Online*, 2:327–344, 2006.

  MS_Align http://atgc.lirmm.fr/

- F. Bonhomme, E. Rivals, A. Orth, G.R. Grant, A. J. Jeffreys, P.R.J. Bois
  Species wide distribution of highly polymorphic minisatellite markers reveals long range gene flow and frequent genetic exchanges among House Mouse subspecies submitted

Credits and collaborators

- ISEM, Montpellier, France
  - F. Bonhomme
  - A. Orth
- Dpt Genetics, Univ. Leicester, GB
  - G. Grant
  - A.J. Jeffreys
- St Jude Childrens Research Hospital, Memphis, USA:
  - P. Bois
- Support: ACI IMPBIO REPEVOL, BioSTIC LR, Génopole LR.

ACI IMPBIO REPEVOL

http://www.lirmm.fr/~rivals/RESEARCH/REPEVOL/