Table 3. Summary of malignancy in patients with anti-aminoacyl-tRNA synthetase antibodies.

Anti-ARS	Age, y	Sex	Diagnosis	ILD	Type of malignancy	Onset
Anti-Jo-1	54	M	PM	−	Lung cancer	At same time
Anti-Jo-1	59	F	DM	+	Gastric cancer	Before DM
Anti-Jo-1	38	F	DM	+	Ovarian cancer	At same time
Anti-Jo-1	54	M	PM	+	Colon cancer	After PM
Anti-Jo-1	74	M	DM	+	Colon cancer	Before DM
Anti-Jo-1	42	F	DM	+	Breast cancer	Before DM
Anti-Jo-1	67	F	DM	+	Non-Hodgkin lymphoma	At same time
Anti-Jo-1	62	M	PM	−	Gastric cancer	After PM
Anti-Jo-1	57	F	DM	+	Thyroid cancer	At same time
Anti-EJ	43	F	DM	+	Nasopharyngeal cancer	At same time
Anti-PL-7	70	F	DM	+	Breast cancer	Before DM
Anti-PL-7	79	M	ILD	+	Gastric cancer	After ILD
Anti-PL-12	53	F	ILD	+	Lung+uterine corpus cancer	Before ILD
Anti-PL-12	66	M	ILD	+	Colon cancer	After ILD
Anti-PL-12	59	F	DM	+	Breast cancer	Before DM
Anti-KS	59	M	ILD	+	Lung cancer	After ILD
Anti-KS	66	M	ILD	+	Prostate cancer	Before ILD
Anti-OJ	71	F	DM	+	Gastric carcinoid	At same time
Anti-OJ	77	M	PM	+	Colon cancer	At same time

ILD: interstitial lung disease; PM: polymyositis; DM: dermatomyositis.
doi:10.1371/journal.pone.0060442.t003

Table 4. Cause of death in patients with anti-aminoacyl-tRNA synthetase antibodies.

Anti-ARS	Age, y	Sex	Diagnosis	ILD	Cause of death	Time after diagnosis (y)
Anti-Jo-1	64	F	DM	+	ILD	0.3
Anti-Jo-1	38	F	DM	+	Infection	3
Anti-Jo-1	36	F	DM	+	ILD	5.5
Anti-Jo-1	62	M	PM	−	Gastric cancer	5
Anti-EJ	65	F	DM	+	ILD	2.5
Anti-EJ	55	F	ILD	+	ILD	0.6
Anti-EJ	55	F	DM	+	ILD	4.25
Anti-EJ	53	F	SSc	+	Infection	6
Anti-EJ	50	F	DM	+	Myocardial infarction	5.25
Anti-PL-7	63	F	DM	+	ILD	1.8
Anti-PL-7	71	F	DM	+	ILD	3
Anti-PL-7	75	M	ILD	+	ILD	0.3
Anti-PL-12	53	F	ILD	+	Lung cancer	3
Anti-PL-12	74	F	DM	+	Rupture of an abdominal aortic aneurysm	0.6
Anti-PL-12	75	F	ILD	+	Hypertrophic cardiomyopathy	2
Anti-KS	59	M	ILD	+	Lung cancer	1.5

ILD: interstitial lung disease; DM: dermatomyositis; PM: polymyositis; SSc: systemic sclerosis.
doi:10.1371/journal.pone.0060442.t004

Table 5. Initial manifestations in patients with anti-aminoacyl-tRNA synthetase antibodies.*

	Anti-Jo-1 (n = 59)	Anti-EJ (n = 38)	Anti-PL-7 (n = 29)	Anti-PL-12 (n = 18)	Anti-KS (n = 13)	Anti-OJ (n = 8)	Overall *P*
DM rashes alone	2	0	14	11	8	0	0.14
Myositis alone	14	11	21	0	0	0	0.14
ILD alone	29	39	28	56	92	63	0.0001[a]
DM rashes and Myositis	10	5	4	6	0	0	0.45
DM rashes and ILD	19	16	10	11	0	0	0.46
Myositis and ILD	7	13	7	0	0	25	0.24
DM rashes, Myositis, and ILD	10	16	17	11	0	13	0.75
No DM rashes, Myositis, or ILD**	10	0	0	6	0	0	0.11

*Values are percentages of patients.
**These patients had polyarthritis at presentation. Significant differences (overall $P<0.05$) were further analyzed by pairwise comparisons.
[a]$P<0.05$ between anti-PL-12 and anti-Jo-1 or anti-KS; $P<0.01$ between anti-KS and anti-Jo-1, anti-EJ or anti-PL-7.
doi:10.1371/journal.pone.0060442.t005

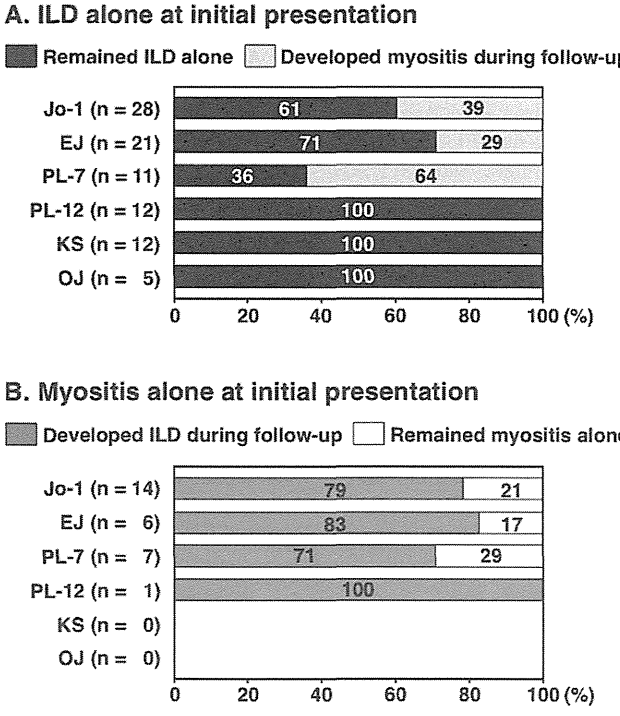

Figure 4. The clinical course of anti-synthetase syndrome patients who developed myositis or interstitial lung disease (ILD) with or without skin manifestations at disease onset. According to the clinical course, patients were classified into four types: remained with ILD alone, developed myositis during follow-up, developed ILD during follow-up, and remained with myositis alone. The clinical course of those who had ILD with or without skin manifestations, but without muscle involvement at their first assessment (A), and the clinical course of those who had myositis with or without skin manifestations, but without ILD at their first assessment (B).
doi:10.1371/journal.pone.0060442.g004

KS, or anti-OJ were less likely to develop myositis during follow-up than those with anti-Jo-1, anti-EJ, or anti-PL-7.

Discussion

This comprehensive report aimed to compare clinical features among anti-ARS-based subgroups on a large scale. As reported previously, more than one anti-ARS Ab did not coexist in general. While this study confirmed that ILD, myositis, Raynaud's phenomenon, polyarthritis, and mechanic's hands were common manifestations in anti-synthetase syndrome, the frequencies of each manifestation varied. That is, myositis was well associated with anti-Jo-1, anti-EJ, and anti-PL-7. Additionally, a substantial number of patients positive for anti-EJ or anti-PL-12 had CADM. Therefore, most of the clinical diagnoses were PM or DM for anti-Jo-1, anti-EJ, and anti-PL-7; CADM or ILD for anti-PL-12; and ILD for anti-KS and anti-OJ. Although patients with anti-ARS Abs share several common manifestations, it is likely that each of these Abs defines a clinically distinct phenotype and may serve as a predictor for clinical complications.

Since nearly all patients with anti-ARS Abs had ILD, this study confirms previous findings that anti-ARS Abs are a marker for ILD [38–42]. Most of the clinical diagnoses in patients with anti-ARS Abs were classic DM, CADM, PM or ILD alone in this study. This finding was also in accordance with previous reports that anti-ARS Abs were highly specific for a proportion of patients with PM, DM, or ILD [4,38,43–45]. However, classic DM, CADM, or PM was found predominantly in patient subgroups with anti-Jo-1, anti-EJ, and anti-PL-7, whereas two-thirds of patients with anti-PL-12 were diagnosed with CADM or ILD. In contrast, anti-KS and anti-OJ were associated with ILD alone. Therefore, it is likely that the clinical diagnosis varies among anti-ARS-based subgroups.

Regarding myositis, it appears that anti-ARS Abs are divided into myositis-related and non-myositis-related subgroups. Anti-Jo-1, anti-EJ, and anti-PL-7 belong to the myositis-related subgroup, since myositis was found in at least half of the patients with these anti-ARS Abs. These findings agreed with previous reports describing a relationship of myositis with anit-Jo-1 [46], anti-EJ [13,17,47,48], and anti-PL-7 [24,49]. In contrast, anti-PL-12, anti-KS, and anti-OJ were not well related to myositis in this study. These results also paralleled those of former reports that anti-KS is highly associated with ILD [32,48]. However, rates of myositis in anti-PL-12 and anti-OJ appear to be different from previous

Table 6. Initial treatment in patients with anti-aminoacyltransfer RNA synthetase antibodies.*

	Anti-Jo-1 (n = 59)	Anti-EJ (n = 38)	Anti-PL-7 (n = 29)	Anti-PL-12 (n = 18)	Anti-KS (n = 13)	Anti-OJ (n = 8)	Overall *P*
No immunosuppressive therapy	7 (4)	5 (2)	3 (1)	11 (2)	38 (5)	13 (1)	0.0070[a]
Initial treatment							
CS oral only	68 (40)	68 (26)	59 (17)	67 (12)	46 (6)	88 (7)	0.45
CS pulse+oral	8 (5)	16 (6)	21 (6)	6 (1)	8 (1)	0 (0)	0.36
CS (pulse and/or oral)+CsA	10 (6)	3 (1)	3 (1)	11 (2)	0 (0)	0 (0)	0.41
CS (pulse and/or oral)+Tac	2 (1)	0 (0)	3 (1)	0 (0)	0 (0)	0 (0)	0.81
CS (pulse and/or oral)+CY (oral and/or iv)	3 (3)	0 (1)	0 (0)	0 (0)	0 (0)	0 (0)	0.82
CS (pulse and/or oral)+CsA or Tac+CY (oral and/or iv)	0 (0)	0 (0)	7 (2)	6 (1)	0 (0)	0 (0)	0.17
CS (pulse and/or oral)+MZR	0 (0)	3 (1)	3 (1)	0 (0)	0 (0)	0 (0)	0.69
CS (pulse and/or oral)+Buc	0 (0)	3 (1)	0 (0)	0 (0)	8 (1)	0 (0)	0.25

*Values are percentages of patients. Patient numbers are given in parenthesis. CS: corticosteroid; CsA: cyclosporine A; Tac: tacrolimus; CY: cyclophosphamide; iv: intravenous administration; MZR: mizoribine; Buc: bucillamine. Significant differences (overall $P<0.05$) were further analyzed by pairwise comparisons.
[a]$P<0.01$ between anti-KS and anti-Jo-1, anti-EJ or anti-PL-7.
doi:10.1371/journal.pone.0060442.t006

reports. Of a total of 47 cases with anti-PL-12, muscle weakness was observed in 27 (57%) patients [16,23,50]. Sato *et al* reported 7 Japanese patients with anti-OJ, in which muscle weakness was seen in 4 patients [51]. Thus, whether anti-PL-12 and anti-OJ are related to myositis remains controversial. Collectively, patients with anti-ARS Abs form a basically homogenous clinical entity, as previously reported; mutual comparisons in this study elucidated certain differences in clinical features among patients with specific anti-ARS Abs.

Regarding skin manifestations, this study revealed an interesting observation. The main clinical diagnoses in anti-Jo-1, anti-EJ, anti-PL-7, and anti-PL-12 were classic DM or CADM. This resulted from the higher frequencies of DM-specific skin manifestations in these patients, which included heliotrope rash and Gottron's signs. However, the distribution of skin manifestations varied among anti-ARS Abs. Only less than 10% of patients with anti-Jo-1 had heliotrope rash, while approximately 20–30% of those with anti-EJ, anti-PL-7, and anti-PL12 had this eruption. On the other hand, the frequency of anti-Jo-1-positive patients who had Gottron's sign was similar compared to those with anti-EJ, anti-PL-7, and anti-PL-12. Thus, the prevalence of DM-specific skin manifestations is not identical among different anti-ARS Abs, even though the main diagnosis is classic DM or CADM.

With respect to the onset of evident manifestations of myositis and ILD, these patients were divided into three groups: i) patients with myositis preceding ILD; ii) patients with ILD preceding myositis; and iii) patients with simultaneous onset of both conditions. We reported previously that the onset of anti-synthetase syndrome is acute, but that the development of myositis may lag behind the onset of ILD in anti-ARS-positive DM patients [38]. A similar finding was described in another report [44]. In this study, most patients with anti-ARS Abs who had myositis without ILD at the onset of disease developed ILD later. On the other hand, the rate of subsequent occurrence of myositis differed among the subsets of anti-ARS Abs when the patients had ILD without myositis as their initial manifestation. Thus, screening and identification of anti-ARS Abs is found to be beneficial in predicting the onset of ILD.

Other than ILD and myositis, previous reports described that arthritis, Raynaud's phenomenon, fever, and mechanic's hands are common clinical features in anti-synthetase syndrome [21,40,44]. There was no significant difference in the frequency of fever in this study. On the other hand, this study revealed some differences in the frequencies of polyarthritis, Raynaud's phenomenon, and mechanic's hands. While these three manifestations were observed with each anti-ARS Ab at a comparable rate, polyarthritis and mechanic's hands were most frequently found with anti-Jo-1, and Raynaud's phenomenon was most frequently found with anti-PL-12. Nonetheless, the differences in frequencies of these manifestations among anti-ARS subgroups were less evident than that with myositis.

We acknowledge several limitations of this study. First, it included a relatively small number of patients with anti-PL-12, anti-KS, or anti-OJ. Second, most facilities enrolled in this study were referral centers. This study had a higher frequency of DM and a relatively lower frequency of PM compared with other similar studies. This may be explained by the fact that our patients were mainly referred to us by rheumatologists, dermatologists, and pulmonologists, and only a few of them were referred by neurologists. Therefore, we cannot exclude selection bias. Third, the possibility cannot be ruled out that coexistence of anti-Ro/SS-A Abs influence the clinical feature of anti-ARS-positive patients with anti-Ro/SS-A Abs, as anti-Ro/SS-A Abs are considered as myositis-associated Abs and form the subgroup. In the analysis of clinical course, possibilities are raised that the short observation period and the differences in treatment potentially affected the results. Additionally, patients who visited to referral centers were examined for the existence of myositis and they were categorized by Bohan and Peter and Sontheimer criteria that are commonly used for diagnosis of myositis in a current condition. However, as clinical features of patients with anti-ARS Abs are largely heterogeneous, it appears difficult to stratify the patients by current criteria. It may be clinically useful to classify the anti-ARS-positive patients based on the type of anti-ARS Abs, not current criteria. It needs to consider the conformity of the classification of the patients with anti-ARS Abs with diagnosis criteria for myositis. Indeed, Connors *et al* have proposed the criteria for anti-ARS syndrome as follows [40]. First, patients must have positive serologic testing for anti-ARS Abs. Then, patients have one or more of the following conditions: Evidence of myositis by Bohan

and Peter criteria, evidence of ILD by American Thoracic Society criteria, evidence of arthritis by clinical examination, radiographic findings, or patient self-report, unexplained, persistent fever, Raynaud's phenomenon, and mechanic's hands. Therefore, more studies are needed for a better general understanding of the clinical characteristics of patients with anti-ARS Abs.

In summary, although anti-ARS Abs share common clinical features, each anti-ARS Ab appears to form some distinct clinical subset. However, the identification of anti-ARS Abs (except for anti-Jo-1) is limited only to certain facilities, as it requires a complicated technique. Establishment of a system routinely available to screen all anti-ARS Abs specificities is needed.

Acknowledgments

We thank Ms. Masako Matsubara, Yuko Yamada, Tomoko Hayashi, and Natsuho Yoshifuji for technical assistance.

Author Contributions

Conceived and designed the experiments: YH MF. Performed the experiments: YH MF. Analyzed the data: YH RY MF. Contributed reagents/materials/analysis tools: YH MF TM K. Kaji K. Komura MH M. Kodera EM KF MS HY SS KT M. Kuwana. Wrote the paper: YH MF M. Kuwana.

References

1. Targoff IN (2002) Laboratory testing in the diagnosis and management of idiopathic inflammatory myopathies. Rheum Dis Clin North Am. 28: 859–890, viii.
2. Mammen AL (2010) Dermatomyositis and polymyositis: Clinical presentation, autoantibodies, and pathogenesis. Ann N Y Acad Sci. 1184: 134–153.
3. Hamaguchi Y, Kuwana M, Hoshino K, Hasegawa M, Kaji K, et al. (2011) Clinical correlations with dermatomyositis-specific autoantibodies in adult Japanese patients with dermatomyositis: a multicenter cross-sectional study. Arch Dermatol. 147: 391–398.
4. Love LA, Leff RL, Fraser DD, Targoff IN, Dalakas M, et al. (1991) A new approach to the classification of idiopathic inflammatory myopathy: myositis-specific autoantibodies define useful homogeneous patient groups. Medicine (Baltimore). 70: 360–374.
5. Mierau R, Dick T, Bartz-Bazzanella P, Keller E, Albert ED, et al. (1996) Strong association of dermatomyositis-specific Mi-2 autoantibodies with a tryptophan at position 9 of the HLA-DR beta chain. Arthritis Rheum. 39: 868–876.
6. Komura K, Fujimoto M, Matsushita T, Kaji K, Kondo M, et al. (2005) Prevalence and clinical characteristics of anti-Mi-2 antibodies in Japanese patients with dermatomyositis. J Dermatol Sci. 40: 215–217.
7. Kaji K, Fujimoto M, Hasegawa M, Kondo M, Saito Y, et al. (2007) Identification of a novel autoantibody reactive with 155 and 140 kDa nuclear proteins in patients with dermatomyositis: an association with malignancy. Rheumatology (Oxford). 46: 25–28.
8. Targoff IN, Mamyrova G, Trieu EP, Perurena O, Koneru B, et al. (2006) A novel autoantibody to a 155-kd protein is associated with dermatomyositis. Arthritis Rheum. 54: 3682–3689.
9. Gunawardena H, Wedderburn LR, North J, Betteridge Z, Dunphy J, et al. (2008) Clinical associations of autoantibodies to a p155/140 kDa doublet protein in juvenile dermatomyositis. Rheumatology (Oxford). 47: 324–328.
10. Fujimoto M, Hamaguchi Y, Kaji K, Matsushita T, Ichimura Y, et al. (2012) Myositis-specific anti-155/140 autoantibodies target transcription intermediary factor 1 family proteins. Arthritis Rheum. 64: 513–522.
11. Sato S, Hirakata M, Kuwana M, Suwa A, Inada S, et al. (2005) Autoantibodies to a 140-kd polypeptide, CADM-140, in Japanese patients with clinically amyopathic dermatomyositis. Arthritis Rheum. 52: 1571–1576.
12. Sato S, Hoshino K, Satoh T, Fujita T, Kawakami Y, et al. (2009) RNA helicase encoded by melanoma differentiation-associated gene 5 is a major autoantigen in patients with clinically amyopathic dermatomyositis: Association with rapidly progressive interstitial lung disease. Arthritis Rheum. 60: 2193–2200.
13. Hirakata M, Mimori T, Akizuki M, Craft J, Hardin JA, et al. (1992) Autoantibodies to small nuclear and cytoplasmic ribonucleoproteins in Japanese patients with inflammatory muscle disease. Arthritis Rheum. 35: 449–456.
14. Yoshida S, Akizuki M, Mimori T, Yamagata H, Inada S, et al. (1983) The precipitating antibody to an acidic nuclear protein antigen, the Jo-1, in connective tissue diseases. A marker for a subset of polymyositis with interstitial pulmonary fibrosis. Arthritis Rheum. 26: 604–611.
15. Mathews MB, Reichlin M, Hughes GR, Bernstein RM (1984) Anti-threonyl-tRNA synthetase, a second myositis-related autoantibody. J Exp Med. 160: 420–434.
16. Bunn CC, Bernstein RM, Mathews MB (1986) Autoantibodies against alanyl-tRNA synthetase and tRNAAla coexist and are associated with myositis. J Exp Med. 163: 1281–1291.
17. Targoff IN (1990) Autoantibodies to aminoacyl-transfer RNA synthetases for isoleucine and glycine. Two additional synthetases are antigenic in myositis. J Immunol. 144: 1737–1743.
18. Hirakata M, Suwa A, Nagai S, Kron MA, Trieu EP, et al. (1999) Anti-KS: identification of autoantibodies to asparaginyl-transfer RNA synthetase associated with interstitial lung disease. J Immunol. 162: 2315–2320.
19. Betteridge Z, Gunawardena H, North J, Slinn J, McHugh N (2007) Anti-synthetase syndrome: a new autoantibody to phenylalanyl transfer RNA synthetase (anti-Zo) associated with polymyositis and interstitial pneumonia. Rheumatology (Oxford). 46: 1005–1008.
20. Hashish L, Trieu EP, Sadanandan P, Targoff IN (2005) Identification of autoantibodies to tyrosyl-tRNA synthetase in dermatomyositis with features consistent with antisynthetase syndrome. Arthritis Rheum 52 Suppl 9: S312.
21. Targoff IN (1992) Autoantibodies in polymyositis. Rheum Dis Clin North Am. 18: 455–482.
22. Hervier B, Devilliers H, Stanciu R, Meyer A, Uzunhan Y, et al. (2012) Hierarchical cluster and survival analyses of antisynthetase syndrome: Phenotype and outcome are correlated with anti-tRNA synthetase antibody specificity. Autoimmun Rev. 12: 210–217.
23. Kalluri M, Sahn SA, Oddis CV, Gharib SL, Christopher-Stine L, et al. (2009) Clinical profile of anti-PL-12 autoantibody. Cohort study and review of the literature. Chest. 135: 1550–1556.
24. Sato S, Hirakata M, Kuwana M, Nakamura K, Suwa A, et al. (2005) Clinical characteristics of Japanese patients with anti-PL-7 (anti-threonyl-tRNA synthetase) autoantibodies. Clin Exp Rheumatol. 23: 609–615.
25. Bohan A, Peter JB (1975) Polymyositis and dermatomyositis (second of two parts). N Engl J Med. 292: 403–407.
26. Sontheimer RD (2002) Would a new name hasten the acceptance of amyopathic dermatomyositis (dermatomyositis sine myositis) as a distinctive subset within the idiopathic inflammatory dermatomyopathies spectrum of clinical illness? J Am Acad Dermatol. 46: 626–636.
27. Tan EM, Cohen AS, Fries JF, Masi AT, McShane DJ, et al. (1982) The 1982 revised criteria for the classification of systemic lupus erythematosus. Arthritis Rheum. 25: 1271–1277.
28. Anonymous (1980) Preliminary criteria for the classification of systemic sclerosis (scleroderma). Subcommittee for Scleroderma Criteria of the American Rheumatism Association Diagnostic and Therapeutic Criteria Committee. Arthritis Rheum. 23: 581–590.
29. Tomasova, Studynkova J, Charvat F, Jarosova K, Vencovsky J (2007) The role of MRI in the assessment of polymyositis and dermatomyositis. Rheumatology (Oxford). 46: 1174–1179.
30. Vitali C, Bombardieri S, Jonsson R, Moutsopoulos HM, Alexander EL, et al. (2002) Classification criteria for Sjogren's syndrome: a revised version of the European criteria proposed by the American-European Consensus Group. Ann Rheum Dis. 61: 554–558.
31. Hamaguchi Y, Hasegawa M, Fujimoto M, Matsushita T, Komura K, et al. (2008) The clinical relevance of serum antinuclear antibodies in Japanese patients with systemic sclerosis. Br J Dermatol. 158: 487–495.
32. Hirakata M, Suwa A, Takada T, Sato S, Nagai S, et al. (2007) Clinical and immunogenetic features of patients with autoantibodies to asparaginyl-transfer RNA synthetase. Arthritis Rheum. 56: 1295–1303.
33. Kuwana M, Kaburaki J, Mimori T, Tojo T, Homma M (1993) Antoantibody reactive with three classes of RNA polymerases in sera from patients with systemic sclerosis. J Clin Invest. 91: 1399–1404.
34. Moroi Y, Peebles C, Fritzler MJ, Steigerwald J, Tan EM (1980) Autoantibody to centromere (kinetochore) in scleroderma sera. Proc Natl Acad Sci U S A. 77: 1627–1631.
35. Vancsa A, Gergely L, Ponyi A, Lakos G, Nemeth J, et al. (2010) Myositis-specific and myositis-associated antibodies in overlap myositis in comparison to primary dermatopolymyositis: Relevance for clinical classification: retrospective study of 169 patients. Joint Bone Spine. 77: 125–130.
36. Bandoh S, Fujita J, Ohtsuki Y, Ueda Y, Hojo S, et al. (2000) Sequential changes of KL-6 in sera of patients with interstitial pneumonia associated with polymyositis/dermatomyositis. Ann Rheum Dis. 59: 257–262.
37. Honda Y, Kuroki Y, Matsuura E, Nagae H, Takahashi H, et al. (1995) Pulmonary surfactant protein D in sera and bronchoalveolar lavage fluids. Am J Respir Crit Care Med. 152: 1860–1866.
38. Matsushita T, Hasegawa M, Fujimoto M, Hamaguchi Y, Komura K, et al. (2007) Clinical evaluation of anti-aminoacyl tRNA synthetase antibodies in Japanese patients with dermatomyositis. J Rheumatol. 34: 1012–1018.
39. Mammen AL (2011) Autoimmune myopathies: autoantibodies, phenotypes and pathogenesis. Nat Rev Neurol. 7: 343–354.
40. Connors GR, Christopher-Stine L, Oddis CV, Danoff SK (2010) Interstitial lung disease associated with the idiopathic inflammatory myopathies: what progress has been made in the past 35 years? Chest. 138: 1464–1474.
41. Gunawardena H, Betteridge ZE, McHugh NJ (2009) Myositis-specific autoantibodies: their clinical and pathogenic significance in disease expression. Rheumatology (Oxford). 48: 607–612.

42. Zong M, Lundberg IE (2011) Pathogenesis, classification and treatment of inflammatory myopathies. Nat Rev Rheumatol. 7: 297–306.
43. Hirakata M, Nagai S (2000) Interstitial lung disease in polymyositis and dermatomyositis. Curr Opin Rheumatol. 12: 501–508.
44. Yoshifuji H, Fujii T, Kobayashi S, Imura Y, Fujita Y, et al. (2006) Anti-aminoacyl-tRNA synthetase antibodies in clinical course prediction of interstitial lung disease complicated with idiopathic inflammatory myopathies. Autoimmunity. 39: 233–241.
45. Katzap E, Barilla-LaBarca ML, Marder G (2011) Antisynthetase syndrome. Curr Rheumatol Rep. 13: 175–181.
46. Mimori T, Imura Y, Nakashima R, Yoshifuji H (2007) Autoantibodies in idiopathic inflammatory myopathy: an update on clinical and pathophysiological significance. Curr Opin Rheumatol. 19: 523–529.
47. Targoff IN, Trieu EP, Plotz PH, Miller FW (1992) Antibodies to glycyl-transfer RNA synthetase in patients with myositis and interstitial lung disease. Arthritis Rheum. 35: 821–830.
48. Watanabe K, Handa T, Tanizawa K, Hosono Y, Taguchi Y, et al. (2011) Detection of antisynthetase syndrome in patients with idiopathic interstitial pneumonias. Respir Med. 105: 1238–1247.
49. Yamasaki Y, Yamada H, Nozaki T, Akaogi J, Nichols C, et al. (2006) Unusually high frequency of autoantibodies to PL-7 associated with milder muscle disease in Japanese patients with polymyositis/dermatomyositis. Arthritis Rheum. 54: 2004–2009.
50. Targoff IN, Arnett FC (1990) Clinical manifestations in patients with antibody to PL-12 antigen (alanyl-tRNA synthetase). Am J Med. 88: 241–251.
51. Sato S, Kuwana M, Hirakata M (2007) Clinical characteristics of Japanese patients with anti-OJ (anti-isoleucyl-tRNA synthetase) autoantibodies. Rheumatology (Oxford). 46: 842–845.

PLOS ONE

Three Groups in the 28 Joints for Rheumatoid Arthritis Synovitis – Analysis Using More than 17,000 Assessments in the KURAMA Database

Chikashi Terao[1,2]*, Motomu Hashimoto[2,3], Keiichi Yamamoto[4], Kosaku Murakami[2], Koichiro Ohmura[2], Ran Nakashima[2], Noriyuki Yamakawa[2], Hajime Yoshifuji[2], Naoichiro Yukawa[2], Daisuke Kawabata[2], Takashi Usui[2], Hiroyuki Yoshitomi[5], Moritoshi Furu[3,5], Ryo Yamada[1,6], Fumihiko Matsuda[1,7,8], Hiromu Ito[3,5], Takao Fujii[2,3], Tsuneyo Mimori[2,3]

1 Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan, **2** Department of Rheumatology and Clinical Immunology, Kyoto University Graduate School of Medicine, Kyoto, Japan, **3** Department of the Control for Rheumatic Diseases, Kyoto University Graduate School of Medicine, Kyoto, Japan, **4** Department of Clinical Trial Design and Management, Translational Research Center, Kyoto University Hospital, Kyoto, Japan, **5** Department of Orthopaedic Surgery, Kyoto University Graduate School of Medicine, Kyoto, Japan, **6** Unit of Statistical Genetics Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan, **7** Institut National de la Sante et de la Recherche Medicale (INSERM) Unite U852, Kyoto University Graduate School of Medicine, Kyoto, Japan, **8** CREST Program, Japan Science and Technology Agency, Kawaguchi, Saitama, Japan

Abstract

Rheumatoid arthritis (RA) is a joint-destructive autoimmune disease. Three composite indices evaluating the same 28 joints are commonly used for the evaluation of RA activity. However, the relationship between, and the frequency of, the joint involvements are still not fully understood. Here, we obtained and analyzed 17,311 assessments for 28 joints in 1,314 patients with RA from 2005 to 2011 from electronic clinical chart templates stored in the KURAMA (Kyoto University Rheumatoid Arthritis Management Alliance) database. Affected rates for swelling and tenderness were assessed for each of the 28 joints and compared between two different sets of RA patients. Correlations of joint symptoms were analyzed for swellings and tenderness using kappa coefficient and eigen vectors by principal component analysis. As a result, we found that joint affected rates greatly varied from joint to joint both for tenderness and swelling for the two sets. Right wrist joint is the most affected joint of the 28 joints. Tenderness and swellings are well correlated in the same joints except for the shoulder joints. Patients with RA tended to demonstrate right-dominant joint involvement and joint destruction. We also found that RA synovitis could be classified into three categories of joints in the correlation analyses: large joints with wrist joints, PIP joints, and MCP joints. Clustering analysis based on distribution of synovitis revealed that patients with RA could be classified into six subgroups. We confirmed the symmetric joint involvement in RA. Our results suggested that RA synovitis can be classified into subgroups and that several different mechanisms may underlie the pathophysiology in RA synovitis.

Citation: Terao C, Hashimoto M, Yamamoto K, Murakami K, Ohmura K, et al. (2013) Three Groups in the 28 Joints for Rheumatoid Arthritis Synovitis – Analysis Using More than 17,000 Assessments in the KURAMA Database. PLoS ONE 8(3): e59341. doi:10.1371/journal.pone.0059341

Editor: Bernhard Kaltenboeck, Auburn University, United States of America

Received October 21, 2012; **Accepted** February 12, 2013; **Published** March 12, 2013

Funding: This study was supported by research grants from Mitsubishi Tanabe Pharma Corporation (http://www.mt-pharma.co.jp/e/), Eisai Co., Ltd. (http://www.eisai.co.jp/index.html), Abbott Japan Co., Ltd. (http://www.abbott.co.jp/), Chugai Pharmaceutical Co., Ltd. (http://www.chugai-pharm.co.jp/hc/ss/english/index.html), Pfizer Japan Inc. (http://www.pfizer.co.jp/pfizer/english/company/), and Bristol-Myyers K.K. (http://www.bms.co.jp/). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. No additional external funding was received for this study.

Competing Interests: The KURAMA database was supported by funding from Mitsubishi Tanabe Pharma Corporation, Eisai Co., Ltd., Abbott Japan Co., Ltd., Chugai Pharmaceutical Co., Ltd., Pfizer Japan Inc. and Bristol-Myyers. This does not alter the authors' adherence to all the PLOS ONE policies on sharing data and materials.

* E-mail: a0001101@kuhp.kyoto-u.ac.jp

Introduction

Rheumatoid arthritis (RA) is the most frequent inflammatory arthritis worldwide affecting 0.5 to 1% of the population [1]. As RA is a bone-destructive disease and functional impairment caused by joint damage is well correlated with swelling and tenderness of joints [2–3], the evaluation of joints in patients with RA is very important to assess disease activity and predict the risk of future joint deformity. ACR core set [4] and DAS (disease activity score) [5–6] were developed for evaluation of disease activity in RA. Recently, the three composite indices, namely, DAS28 [5], simplified disease activity index (SDAI) [7] and clinical disease activity index (CDAI) [8] are frequently used for disease activity evaluation among rheumatologists. All of the three indices are shown to be well correlated with future joint destruction [7,9]. These three methods include the same 28 joints for evaluation of disease activity, namely, bilateral wrist, 1st to 5th metacarpal (MCP) joints and proximal interphalangeal (PIP) joints, elbow, shoulder, and knee joints. Though RA is known to show symmetric joint symptoms [10], the frequency of bilateral joint symptoms and the correlations between each joint symptom are not fully analyzed by using large numbers of joint assessments. There are several reports of successful prediction of joint damage using a reduced number of joints for evaluation by ultrasonogra-

phy [11–12]. These reports raise the possibility that some of the 28 joints are less frequently involved, and are less informative for disease activity. Analyses for characterization of joint symptoms would uncover correlations of unexpected joint symptoms and distribution of synovitis in RA.

Here, we analyzed the distribution of affected joints in the 28 joints in patients with RA using more than 17,000 joint assessments from 1,314 patients with RA and showed that synovitis in RA patients can be classified into three groups. We also showed that affected rates of the 28 joints greatly vary in RA patients, and that RA patients could be classified into subgroups based on the distribution of joint synovitis.

Results

Frequency order of joints involvement

We recruited 17,311 assessments for the 28 joints in 1,314 patients with RA from 2005 to 2011. A summary of the registered patients is listed in Table 1. The distribution of the number of patients with RA in each year and the number of joint assessments for each patient are shown in Figure S1. We analyzed how often each of the 28 joints was tender or swollen in patients with RA in 2011. From the analysis of 735 patients, we found that the frequency of joint swelling and tenderness in the 28 joints is widely different from joint to joint (Figure 1 and Table S1). The wrist joints were the most frequently affected joints for swelling and tenderness. The frequency of the right wrist joint being affected was more than four times as high as the least frequently affected joint. Many of the joints showed right-dominant tenderness (eleven of fourteen joints, $p = 0.057$, binomial test), indicating mostly right-handedness. We found strong correlations for the affected rates of each joint between swellings and tenderness except for shoulder joints (Spearman's rank-sum coefficient, $rho = 0.70$ and $p = 3.8 \times 10^{-5}$, Figure 1, Table S1). Shoulder joints showed much higher frequencies of tenderness than those of swellings.

Next, we tried to replicate the order of affected frequencies of the 28 joints and the correlation between tenderness and swellings in different RA patients. We obtained 579 patients whose joints data were not available for 2011, indicating we analyzed independent RA patients. We found that the order of the affected joint frequencies were well correlated for both swelling and tenderness among different sets of RA patients (Spearman's rank-sum coefficient, rho:0.815 and 0.904, $p = 1.3 \times 10^{-7}$ and $p = 4.6 \times 10^{-11}$ for swelling and tenderness, respectively, Figure S2). We also confirmed that rates of tenderness were well correlated with those of swellings in the 28 joints in the 579 patients (rho:0.604). These results indicate that some of the 28 joints are more likely to develop arthritis than the others in RA patients. The swelling and tenderness correlate with each other except for shoulder joints.

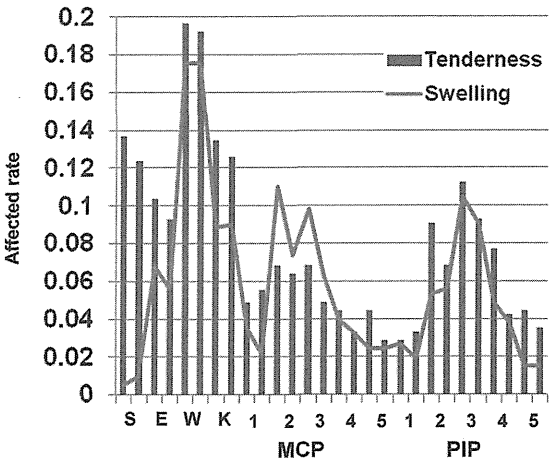

Figure 1. Affected rate of joint symptoms. Affected rate of joint symptoms. Each joint is arranged in the order of right and left. S:shoulder, E:elbow, W:wrist, K:knee.
doi:10.1371/journal.pone.0059341.g001

Table 1. Summary of the KURAMA database.

	The KURAMA database
RA patients	1314
Age (mean±SD)	60.2±15.1
female ratio	81.70%
disease duration (years)	12.2±9.8
Stage*	2.75±1.17
Class*	1.87±0.69

*Stage and Class represent Steinbrocker's stage and class, respectively.
SD: standard deviation.
doi:10.1371/journal.pone.0059341.t001

Whether the right-dominant involvement of joints in patients with RA is associated with joint destruction was analyzed. Joint destruction in the hand was evaluated for 246 patients with RA by modified Sharp score [13]. The six elements of the scores were separately analyzed, namely erosion of PIP, MCP, and wrist joints (we defined as joints other than MCP and PIP in hand) and narrowing of PIP, MCP, and wrist joints. We found that five out of six elements showed right-dominant destruction. In particular, narrowing and erosion of MCP joints showed a statistically significant right-dominance in binomial test ($p <= 0.0050$, Table S2).

Three groups of 28 joints in RA synovitis

Next we analyzed correlations of joint symptoms between the 28 joints. We randomly picked up one assessment from each of the 1,314 patients to maximize the power. When the correlation of tenderness of the 28 joints was analyzed with kappa coefficient, we confirmed that each joint showed a symmetric involvement (Figure 2A). The results also showed that the tenderness of large joints and wrist joints are not correlated with the tenderness of PIP and MCP joints. We found that the tenderness of MCP joints was especially well correlated with each other and that PIP joints tenderness was well correlated with each other. The correlation of swelling in the 28 joints showed the same tendency as that of tenderness, namely, symmetric joint involvement, correlations between large joints and wrist joints, and no strong correlations between wrist joints and other small joints (Figure 2B).

Next we used eigen vectors of principal component analysis to assess the correlations of the 28 joints involvement. When we analyzed correlations of tenderness, eigen vectors revealed that PIP and MCP joints can be clearly distinguished from large joints and wrist joints (Figure 3A). PIP joints and MCP joints turned out to make independent groups after excluding large joints and wrist joints (Figure 3B). These three groups of affected joints were found both for tenderness and swelling (Figure 3C and 3D). We confirmed these three correlation groups in four independent resampling analyses by randomly picking up one assessment from each of the 1,314 patients four times (data not shown). The three groups were observed in the two independent sets of RA patients which were used in the analysis of joints involvement frequency

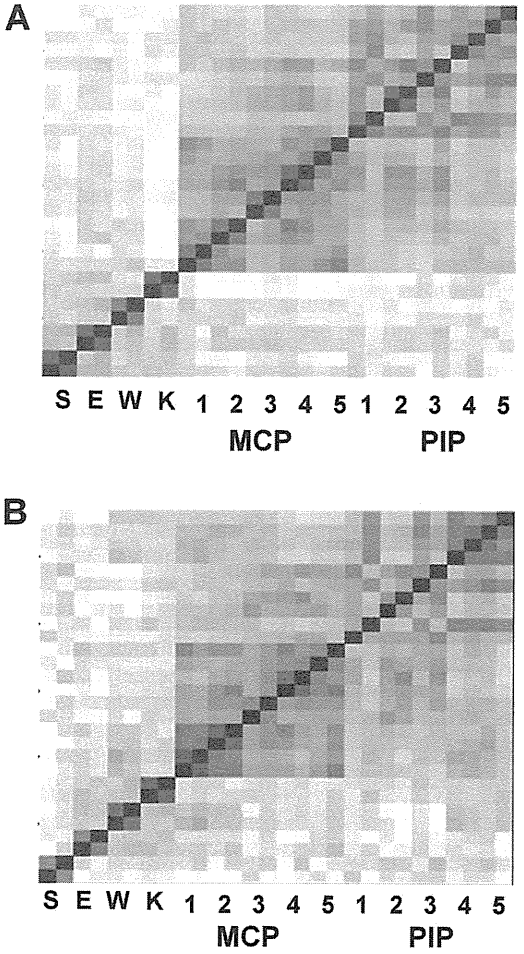

Figure 2. Correlations between the 28 joint symptoms. Brightness of the red color corresponds to the strength of correlations between joint tenderness (A) or swellings (B), using the Kappa coefficient. Each joint is arranged in the order of right and left. The joint order in the y axis is the same as the x axis. The result is a representative of five analyses based on resampled assessments. S:shoulder, E:elbow, W:wrist, K:knee.
doi:10.1371/journal.pone.0059341.g002

(Figure S3). In addition, no significant difference was observed in the relationship of the three groups of joint involvement when we divided the 1,314 patients into two groups according to the patients' caring physicians (Figure S4). We confirmed the three groups by resampling four times for each analysis (data not shown). These results indicate that these three groups were not due to specific patients, examiners, or time of evaluation.

Taken together, the correlation analyses using kappa coefficient and eigen vectors in principal component analysis indicated that there are three correlated groups of joints in RA synovitis, namely, large joints with wrist joints (which we express as "large and wrist joints"), PIP joints, and MCP joints.

Subgroups of patients with RA

We performed a clustering analysis of 5,383 evaluations of 28 joints from 1,314 patients with RA. Six subgroups of evaluations of 28 joints were observed (Figure 4). Each of the subgroups was characterized by 1) no synovitis (34.6%), 2) mild activity with dominant involvement of large and wrist joints (17.4%), 3) dominant involvement of MCP joints (18.3%), 4) dominant involvement of PIP joints (9.3%), 5) active synovitis (4.1%), and 6) moderate activity with dominant involvement of large and wrist joints (16.4%) (Table S3). Whether patients with RA are classified into the same subgroups was analyzed. There were 998 patients with four or five evaluations, and of these, 734 were categorized into the regular groups across different evaluations, indicating that the patterns of synovitis in the same patients were stable. Analysis of joint destruction in each subgroup revealed that the sixth subgroup demonstrated dominant destruction of large and wrist joints compared with MCP and PIP joints ($p <= 2.8 \times 10^{-5}$, Figure S5 and Figure S6).

Discussion

Since RA is a joint destructive autoimmune arthritis and joint damage occurs rapidly in the early stages of the disease course [14], the development of a quantitative scale which assesses disease activity and predicts joint damage is very important. After DAS and ACR core sets were introduced, DAS28, SDAI, and CDAI were developed to evaluate disease activity and easily calculate the disease activity score in patients with RA. All three indices were shown to be well correlated with future joint destruction and they share the same 28 joints for evaluation. Joint symptoms especially joint swelling is known to correlate with future joint damage [3]. While these indices were developed for use in clinical trials such as responsiveness to treatment, they are used by rheumatologists in daily clinical practice and they are reported to coincide very well among different examiners [9]. Characterizing the relative affected frequency of each joint and analysis of correlation between joint symptoms are important to analyze the basic mechanisms of synovitis and to efficiently select the joints to predict future joint destruction. However, there is no detailed analysis to address the correlations between the 28-joint symptoms.

In the current study, we characterized the 28-joint symptoms using large numbers of joint assessments. While we reported the affected rates of each joint in the 28 joints for tenderness and swelling of RA patients registered in the KURAMA database in 2011 as a representative (Table S1), these rates should not be generalized considering large effects of treatment especially biologics agents on joint symptoms. Thus, we focused on relative frequencies of joint involvement for the 28 joints. The affected frequency pattern was compared between the two sets of RA patients, and there were no apparent differences between the two sets for both tenderness and swelling. We also showed that joint symptoms in RA could be classified into three groups both for tenderness and swelling. Our analysis also demonstrated that patients with RA can be regularly classified into six subgroups based on patterns of joint symptoms. These results suggest that regular RA joint involvement pattern, including relative frequency and groups of joints, is largely maintained in RA patients. In addition, we confirmed that these patterns of joint involvement were not attributed to evaluators and fractions of RA patients.

It is interesting that the affected frequencies greatly varied from joint to joint, and the rate of the most highly affected joint was more than four times as high as the least-affected joint. The affected frequencies indicated that wrist joints were the most frequently affected. It should be noted that surface area may have influenced the sensitivity of detecting synovitis in physical exams when different joints were compared. The relatively high frequency of tenderness and swelling in large and wrist joints compared with MCP and PIP joints can be explained by this difference in surface area. However, surface area cannot fully explain the highest frequency of wrist involvement and different frequencies within the MCP or PIP joints. A dominant involve-

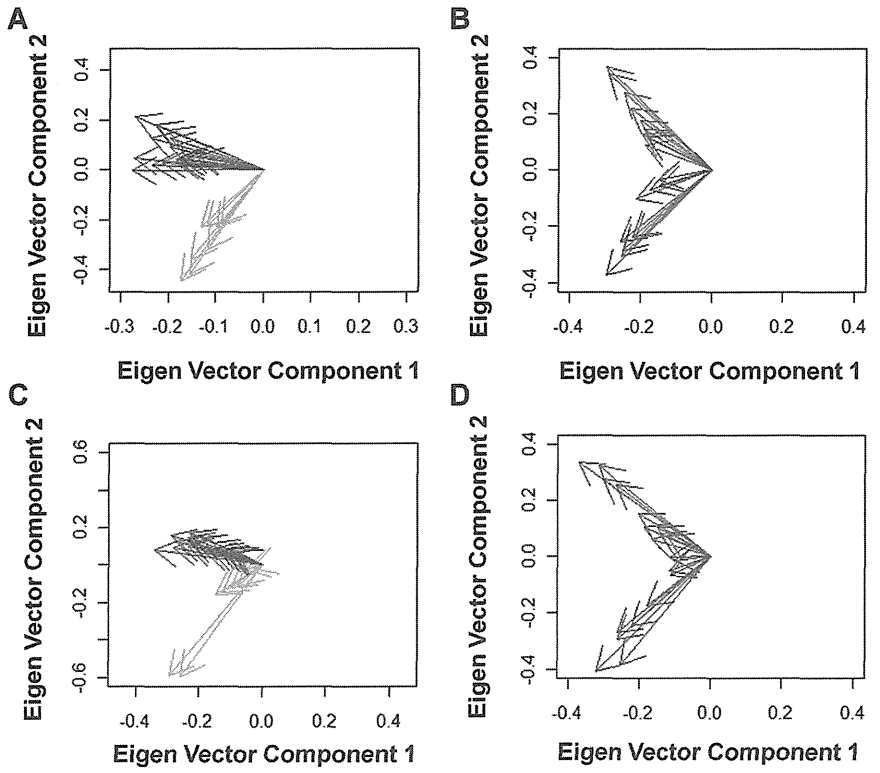

Figure 3. Relationship of the 28-joint involvement. The 1st and 2nd components of eigen vectors of the joint symptoms are plotted, using principal component analysis of the 28 joint involvement for tenderness (A) and swelling (C) or using that of the 20 joint involvement other than large and wrist joints for tenderness (B) and swelling (D). The results are representatives of five analyses based on resampled assessments. Green: large and wrist joints. Red: MCP joints. Blue: PIP joints.
doi:10.1371/journal.pone.0059341.g003

ment of right joints seemed to indicate a majority of the study population being right-handed in spite of the small difference of affected rates between bilateral joints. We also demonstrated that the right dominant involvement was also true for joint destruction. We could not compare the joint involvement and joint destruction between right-handed patients and left-handed patients due to a lack of information regarding handedness of patients.

Correlation analysis confirmed the well-known symmetric joint involvement in patients with RA. Strong correlations of tenderness and swelling in the same joints except for shoulder joints may indicate low sensitivity of shoulder swelling in the physical exams and common mechanisms of swelling and tenderness. It is striking that joint symptoms can be classified into three groups based on correlation analysis and principal component analysis. The association observed between the symptoms in the wrist joints and the large joints is worth noting, since wrist joints are regarded as small joints according to ACR/EULAR criteria set in 2010. As wrist joints are much closer to other small joints than large joints, the relationship between wrist joints and large joints cannot be explained by the distance of joints. The distance of joints cannot explain the two different groups of MCP and PIP joints either. While symptoms of large and wrist joints are not related with those of MCP and PIP joints, they were not very strongly correlated with each other, compared with correlations among PIP joints or MCP joints. This may indicate that there are no common strong factors which predispose large and wrist joints to swelling and tenderness in patients with RA.

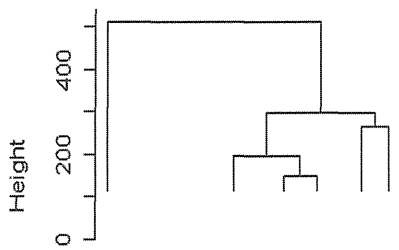

Figure 4. Six subgroups of evaluations of the 28 joints in RA. Results of clustering analysis with Ward method using randomly obtained 5,383 evaluations of the 28 joints in 1,314 patients were plotted.
doi:10.1371/journal.pone.0059341.g004

We also showed that patients with RA can be divided into six subgroups based on these three groups of joint involvement. More than 70% of patients are classified into regular subgroups, indicating that the pattern of synovitis in a patient with RA is stable. When patients who were regularly classified into the first subgroup of patients characterized by no synovitis were removed, more than 60% of patients were still classified into regular subgroups (data not shown), indicating that the stable patterns were observed regardless of activity of RA. As joint destruction was influenced by disease duration, disease activity, and treatment, we analyzed the relative distribution of joint destruction between the three joint groups in a patient with RA. We found that the sixth subgroup of patients, characterized by moderate activity with dominant involvement of large and wrist joints, demonstrated dominant destruction of wrist joints. This suggests that classifying patients with RA into appropriate subgroups would lead to prediction of patterns of joint destruction.

There are reports that evaluating fraction of joints by ultrasonography is a good way to predict future joint damage [11–12]. One study reported that 5 of the 28 joints with MTP2 and MTP5 joints, namely, wrist, MCP2, MCP3, PIP2, and PIP3 joints, are enough for ultrasonography evaluation [12]. Their data seems to be consistent with our results as they selected at least two joints from three different groups into which the 28-joint symptoms were classified. As ultrasonography usually surpasses physical examination in terms of the sensitivity to detect synovitis, it is interesting to analyze whether the assessments of synovitis using ultrasonography show the same pattern of synovitis over the 28 joints in RA.

Our results indicate that RA does not develop synovitis in the 28 joints with the same frequency and that the affected rate of each joint greatly varies from joint to joint. These different distributions of joint synovitis would lead to different distribution of joint destruction. Based on our results, the 28 joints can be categorized into three groups, and it is possible that some fractions of the 28 joints are less informative to assess disease activity than others. It would be interesting to develop a novel simplified joint core set, and analyze the correlation between joint damage and activity score based on this. It would be also interesting to characterize each of RA subsets in more detail.

Materials and Methods

Ethics Statement

Written informed consent to enroll in the database described below was obtained from most of the patients, but for some patients the information regarding the construction of this database was disclosed instead of obtaining written informed consent. Participants who were informed regarding the construction of the database (instead of obtaining written informed consent) were allowed to withdraw from the study if desired.

All data were de-identified and analyzed anonymously. This study was designed in accordance with the Helsinki Declaration. This study including the consent procedure was approved by the ethics committee of Kyoto University Graduate School and Faculty of Medicine.

The KURAMA database

The KURAMA (Kyoto University Rheumatoid Arthritis Management Alliance) database was established in 2011 at Kyoto University to store detailed clinical information and specimens from patients with arthritis and arthropathy. The alliance is composed of rheumatic disease-associated departments in Kyoto University Hospital as well as its allied, integrating previous database and specimen collections in each department and allied. A template for electronic clinical charts developed at Kyoto University Hospital in 2004 to evaluate joint involvements in RA patients was used to obtain joint assessments. Rheumatologists evaluated swelling and tenderness of the 28 joints in patients with RA on each visit and filled in the template. The synovitis information of the 28 joints and data for C-reactive protein and erythrocyte sedimentation rate were extracted from electronic clinical charts [15] and stored in the KURAMA database.

Patients and data of joint assessment

A total of 17,311 joint assessments from 1,314 patients with RA from 2005 to 2011 were obtained in a retrospective manner from the KURAMA database. All of the patients fulfilled ACR revised criteria for RA in 1987 [10] or ACR and EULAR classification criteria for RA in 2010 [16–17].

Analysis of affected frequencies in the 28 joints

RA patients were subdivided depending on whether their data were available in 2011 or not, and the affected frequency in each of the 28 joints was calculated. We compared the order of the affected frequency in the 28 joints between the two patient sets with Spearman's rank-sum coefficient. We separately analyzed the affected rates of joints for swelling and tenderness. When multiple joint assessments in different visits were available in the same patient with RA, we randomly selected one of the assessments as representative in the patient. We compared frequencies between tenderness and swellings for the 28 joints with Spearman's rank-sum coefficient.

Clustering of patients with RA

Clustering analyses were performed by Ward method, using randomly-selected 5,383 evaluations of the 28 joints from 1,314 patients with RA. These evaluations did not contain more than six assessments from each patient to avoid excess influence of particular patients. Affected rates were calculated for the three groups of joints (namely PIP joints, MCP joints and large and wrist joints) in this clustering analysis. For example, when a patient showed tenderness and swelling for all PIP joints, the affected rate of PIP joints in the patient is 2. When a patient showed tenderness for four MCP joints, the affected rate of MCP joints is 0.4.

RA patients were regarded as belonging to a particular group when more than 60% of evaluations belonging to the same patients with four or five evaluations were classified into the same group.

Analysis between RA subgroups and joint destruction

Joint destruction of hand joints in 246 patients with RA was evaluated by modified Sharp score by a trained rheumatologist who was not informed of the patients' characteristics (KM). Joint destruction rates were defined for the three groups of joints as a sum of scores divided by the full score in the joints group. For example, when a patient shows 50 as a sum of scores in the large and wrist group, the patient's joint destruction rate for the group is 0.463 (50/108).

Correlation of the 28 joints and statistical analysis

Correlations of joint symptoms among the 28 joints were estimated separately for tenderness and swelling. We randomly obtained one assessment of the 28 joints in each patient as a representative of the patient's joint assessments for maximization of the power. Kappa coefficient was used to analyze coincidence of joint symptoms in each pair of the 28 joints. Eigen vectors obtained in principal component analysis were used to analyze the deviation of joint symptoms. We resampled joint assessments for each patient and created four other sets of joint assessments. The same correlation analyses were performed using the four resampled assessments to confirm the correlation shown in the first assessment set. Right dominance of the synovitis and joint destruction was analyzed by binomial test. Dominant destruction of joints was evaluated by paired-t test. Statistical analysis was performed by R software or SPSS (ver18).

Supporting Information

Figure S1 Distribution of joint evaluation counts and patients across different years. A) Distribution of number of RA patients according to numbers of 28-joint assessments. B) Distribution of number of patients with RA whose joint assessment data were available from 2005 to 2011 in the KURAMA database. (TIF)

Figure S2 Good correlations between joint involvement rates in different sets of RA patients. Rates of joint involvement for A) swelling and B) tenderness were compared between the two different sets of RA patients. X and Y axes represent rates in the first set of RA patients in 2011 and those in the second set in 2005 to 2010, respectively.
(TIF)

Figure S3 Three groups of joints regardless of different sets of RA patients. Analysis using one of four resampled assessments in one of the two sets of RA patients is shown as a representative. The 1st and 2nd components of eigen vectors of the joint symptoms are plotted, using principal component analysis of the 28 joint involvement for tenderness (A) and swelling (C) or using that of the 20 joint involvement other than large and wrist joints for tenderness (B) and swelling (D). Green: large and wrist joints. Red: MCP joints. Blue: PIP joints.
(TIF)

Figure S4 Three groups of joints regardless of different evaluators. Analysis using one of five resampled assessments by one of the two groups of medical doctors is shown as a representative. The 1st and 2nd components of eigen vectors of the joint symptoms are plotted, using principal component analysis of the 28 joint involvement for tenderness (A) and swelling (C) or using that of the 20 joint involvement other than large and wrist joints for tenderness (B) and swelling (D). Green: large and wrist joints. Red: MCP joints. Blue: PIP joints.
(TIF)

Figure S5 Dominant destruction of large and wrist joints in the sixth subgroup of patients with RA. Box plots indicating the joint destruction rates in the three joint groups in subjects belonging to the sixth subgroup.
(TIF)

Figure S6 Destruction of large and wrist joints among the six subgroups of RA. Differences in destruction rates were plotted for each subject in the six subgroups. The difference was defined as: A) destruction rate of group of large and wrist joints – destruction rate of MCP joints and B) destruction rate of group of large and wrist joints – destruction rate of PIP joints.
(TIF)

Table S1 Rate of joint involvement for 28 joints in RA.
(DOC)

Table S2 Right-dominant joint destruction in RA. Patients who showed unilateral higher or lower scores in each element were analyzed.
(DOC)

Table S3 Mean affected rates of the three joint groups in the six subgroups of patients with RA.
(DOC)

Acknowledgments

We would like to thank to Mr. Wataru Yamamoto at Kurashiki Kosai Hospital for his excellent support to establish and maintain the KURAMA database. We also thank Drs Hisashi Yamanaka, Katsunori Ikari, and Ayako Nakajima at Institute of Rheumatology, Tokyo Women's Medical University for their kind instruction and advice for management of rheumatic diseases database.

Author Contributions

Evaluation of joint X-rays: KM. Conceived and designed the experiments: CT MH KO RY FM HI TF TM. Analyzed the data: CT. Contributed reagents/materials/analysis tools: CT MH KO RN KM N. Yamakawa H. Yoshifuji N. Yukawa DK TU H. Yoshitomi MF HI TF TM KY. Wrote the paper: CT.

References

1. Firestein GS (2003) Evolving concepts of rheumatoid arthritis. Nature 423: 356–361.
2. Drossaers-Bakker KW, de Buck M, van Zeben D, Zwinderman AH, Breedveld FC, et al. (1999) Long-term course and outcome of functional capacity in rheumatoid arthritis: the effect of disease activity and radiologic damage over time. Arthritis and Rheumatism 42: 1854–1860.
3. Smolen JS, Van Der Heijde DM, St Clair EW, Emery P, Bathon JM, et al. (2006) Predictors of joint damage in patients with early rheumatoid arthritis treated with high-dose methotrexate with or without concomitant infliximab: results from the ASPIRE trial. Arthritis and Rheumatism 54: 702–710.
4. Felson DT, Anderson JJ, Boers M, Bombardier C, Chernoff M, et al. (1993) The American College of Rheumatology preliminary core set of disease activity measures for rheumatoid arthritis clinical trials. The Committee on Outcome Measures in Rheumatoid Arthritis Clinical Trials. Arthritis and Rheumatism 36: 729–740.
5. van der Heijde DM, van 't Hof MA, van Riel PL, Theunisse LA, Lubberts EW, et al. (1990) Judging disease activity in clinical practice in rheumatoid arthritis: first step in the development of a disease activity score. Annals of the Rheumatic Diseases 49: 916–920.
6. van der Heijde DM, van't Hof MA, van Riel PL, van Leeuwen MA, van Rijswijk MH, et al. (1992) Validity of single variables and composite indices for measuring disease activity in rheumatoid arthritis. Annals of the Rheumatic Diseases 51: 177–181.
7. Smolen JS, Breedveld FC, Schiff MH, Kalden JR, Emery P, et al. (2003) A simplified disease activity index for rheumatoid arthritis for use in clinical practice. Rheumatology 42: 244–257.
8. Aletaha D, Smolen JS (2007) The Simplified Disease Activity Index (SDAI) and Clinical Disease Activity Index (CDAI) to monitor patients in standard clinical care. Best Pract Res Clin Rheumatol 21: 663–675.
9. Salaffi F, Cimmino MA, Leardini G, Gasparini S, Grassi W (2009) Disease activity assessment of rheumatoid arthritis in daily practice: validity, internal consistency, reliability and congruency of the Disease Activity Score including 28 joints (DAS28) compared with the Clinical Disease Activity Index (CDAI). Clinical and Experimental Rheumatology 27: 552–559.
10. Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, et al. (1988) The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum 31: 315–324.
11. Scheel AK, Hermann KG, Kahler E, Pasewaldt D, Fritz J, et al. (2005) A novel ultrasonographic synovitis scoring system suitable for analyzing finger joint inflammation in rheumatoid arthritis. Arthritis and Rheumatism 52: 733–743.
12. Backhaus M, Ohrndorf S, Kellner H, Strunk J, Backhaus TM, et al. (2009) Evaluation of a novel 7-joint ultrasound score in daily rheumatologic practice: a pilot project. Arthritis and Rheumatism 61: 1194–1201.
13. van der Heijde D (2000) How to read radiographs according to the Sharp/van der Heijde method. Journal of Rheumatology 27: 261–263.
14. Machold KP, Stamm TA, Eberl GJ, Nell VK, Dunky A, et al. (2002) Very recent onset arthritis – clinical, laboratory, and radiological findings during the first year of disease. Journal of Rheumatology 29: 2278–2287.
15. Yamamoto K, Yamanaka K, Hatano E, Sumi E, Ishii T, et al. (2012) An eClinical trial system for cancer that integrates with clinical pathways and electronic medical records. Clin Trials 9: 408–417.
16. Aletaha D, Neogi T, Silman AJ, Funovits J, Felson DT, et al. (2010) 2010 Rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. Arthritis and Rheumatism 62: 2569–2581.
17. Aletaha D, Neogi T, Silman AJ, Funovits J, Felson DT, et al. (2010) 2010 rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. Annals of the Rheumatic Diseases 69: 1580–1588.

Gene 505 (2012) 324–332

Contents lists available at SciVerse ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Extremely slow rate of evolution in the HOX cluster revealed by comparison between Tanzanian and Indonesian coelacanths

Koichiro Higasa [a,1], Masato Nikaido [b,1], Taro L. Saito [a,1], Jun Yoshimura [a,1], Yutaka Suzuki [c], Hikoyu Suzuki [b], Hidenori Nishihara [b], Mitsuto Aibara [b], Benjamin P. Ngatunga [d], Hassan W.J. Kalombo [e], Sumio Sugano [c], Shinichi Morishita [a,*], Norihiro Okada [b,**]

[a] Department of Computational Biology, Graduate School of Frontier Sciences, The University of Tokyo, 5-1-15 Kashiwanoha, Kashiwa City, Chiba 277-0882, Japan
[b] Department of Biological Sciences, Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, 4259-B21 Nagatsuta-cho, Midori-ku, Yokohama 226-8501, Japan
[c] Department of Medical Genome Sciences, Graduate School of Frontier Sciences, The University of Tokyo, 5-1-15 Kashiwanoha, Kashiwa City, Chiba 277-0882, Japan
[d] Tanzania Fisheries Research Institute, P.O. Box 9750, Dar es Salaam, Tanzania
[e] Regional Commissioner's Office Tanga, Box 5095, Tanga, Tanzania

ARTICLE INFO

Article history:
Accepted 21 May 2012
Available online 12 June 2012

Keywords:
Living fossil
Synonymous substitution rate
Next generation sequencer

ABSTRACT

Coelacanths are known as "living fossils" because their morphology has changed very little from that in the fossil record. To elucidate why coelacanths have evolved so slowly is thus of primary importance in evolutionary biology. In the present study, we determined the entire sequence of the HOX cluster of the Tanzanian coelacanth (*Latimeria chalumnae*) and compared it with that of the Indonesian coelacanth (*L. menadoensis*), which was available in the literature. The most intriguing result was the extremely small genetic divergence between the two coelacanths. The synonymous divergence of the HOX coding region between the two coelacanths was estimated to be 0.07%, which is ~11-fold smaller than that of human–chimp. When we applied the estimated divergence time of the two coelacanths of 6 million years ago (MYA) and 30 MYA, which were proposed in independent mitochondrial DNA analyses, the synonymous substitution rate of the coelacanth HOX cluster was estimated to be ~11-fold and 56-fold smaller than that of human–chimp, respectively. Thus, the present study implies that the reduction of the nucleotide substitution rate in coelacanth *HOX* genes may account for the conservation of coelacanth morphology during evolution.

1. Introduction

Coelacanths were initially recognized as a distinct taxonomic group by fossil records, which range in age from the Early Devonian to the Late Cretaceous periods (Maisey, 1996). Because no fossil records of coelacanths have been found since 65 million years ago (MYA), coelacanths were believed to have been extinct (Maisey, 1996). Therefore, the discovery in 1938 of the first living coelacanth, *Latimeria chalumnae*, off the coast of South Africa created a sensation in the field of evolutionary biology (Smith, 1939). One of the most interesting observations is that coelacanth morphology has changed very little (Holder et al., 1999; Smith, 1939); most of the characteristics unique to coelacanths (fleshy-lobed fins, hollow nerve cord, poor ossification of the skeleton, lack of defined ribs, and bilobed caudal region) have been maintained from the Devonian era (Carroll, 1988). Accordingly, coelacanths are called "evolutionary relics" or "living fossils". The elucidation of the molecular mechanism of such slow morphological change in coelacanths is of primary importance to understand the morphological evolution of animals from genotype to phenotype.

After the discovery of a second living coelacanth in the Comoros archipelagos (Smith, 1953), the existence of a viable coelacanth population in this region was confirmed. In addition to the Comoros archipelagos, several coelacanths have been captured off the coasts of Mozambique (Schliewen et al., 1993), Madagascar (Heemstra et al., 1996), and Kenya (De Vos and Oyugi, 2002). Nikaido et al. (2011) recently found a genetically distinct coelacanth population off the northern coastal region of Tanzania, indicating that coelacanths are widely distributed throughout the western Indian Ocean. Apart from the western Indian Ocean, two coelacanth individuals were captured off the coast of Manado, Sulawesi, Indonesia (Erdman et al., 1998). These coelacanths are the first individuals recorded from a location outside the western Indian Ocean and were described as a new species, *Latimeria menadoensis*. The divergence time between the two coelacanth species was estimated in three independent studies. Holder et al. (1999) used partial mtDNA sequences and estimated the divergence time at 6 MYA. On the other hand, using the entire mtDNA sequences

Abbreviations: MYA, million years ago; PCR, polymerase chain reaction; SNV, single nucleotide variation; SINE, short interspersed repetitive element.

* Corresponding author. Tel.: +81 47 136 3985; fax: +81 47 136 3977.
** Corresponding author. Tel.: +81 45 924 5742; fax: +81 45 924 5835.
E-mail addresses: moris.utokyo@gmail.com (S. Morishita), nokada@bio.titech.ac.jp (N. Okada).

[1] These authors equally contributed to this work.

0378-1119/$ – see front matter.
doi:10.1016/j.gene.2012.05.047

(except for the d-loop) and Bayesian methods, Inoue et al. (2005) estimated the divergence time at about 30 MYA. Sudarto et al. (2010) also used Bayesian analysis and proposed the divergence time to be 28 MYA.

HOX genes encode a highly conserved family of transcription factors possessing a 60-amino acid residue motif called a homeodomain, and they are involved in morphogenesis during embryonic development (Krumlauf, 1994). Most of the jawed vertebrates have four separate HOX clusters–HOXA, HOXB, HOXC and HOXD–in which about 40 *HOX* genes are arranged. However, the number and composition of HOX clusters of teleost fish are distinct from those of the other vertebrates owing to a whole-genome duplication event specific to teleost fish (Meyer and Málaga-Trillo, 1999). In particular, duplication of the HOX clusters led to eight HOX clusters in an ancestor of teleost fish. Subsequently, some *HOX* genes were lost independently in the lineage of each teleost fish species during evolution. As a result, euteleosts have seven HOX clusters, in which 46 (medaka) to 49 (zebrafish) *HOX* genes were identified (Kurosawa et al., 2006). Amemiya et al. (2010) characterized the complete HOX clusters in the coelacanth genome. Although the HOX clusters of coelacanth were not remarkable relative to those from other species with four clusters, characterization of the complete *HOX* genes of coelacanth allowed us to reconstruct the evolutionary history of HOX clusters among vertebrates.

Consistent with the slow rate of phenotypic changes in coelacanth, several genetic studies showed a slow rate of evolution at the molecular level. Noonan et al. (2004) indicated that the content and organization of the procadherin gene cluster were more conserved in coelacanths than in teleost fish. Furthermore, they indicated fewer amino acid substitutions in coelacanths than in zebrafish and humans. Amemiya et al. (2010) also showed a significantly slower rate of amino acid substitution in the coelacanth HOX cluster than in teleost fish and tetrapods. These observations suggested that the coelacanth genes have evolved under strong purifying selection or that mutation rate has been slowed down in the coelacanth genome. To examine these possibilities, we directly estimated the absolute value of synonymous divergence of the two coelacanths. Namely, we determined the entire HOX cluster sequence for Tanzanian coelacanth *L. chalumnae* and compared it to that of *L. menadoensis*, which was already available in the literature (Amemiya et al., 2010).

2. Materials and methods

2.1. Genomic DNA of Tanzanian coelacanth

Frozen or ethanol-preserved coelacanth materials were transferred from the Tanzania Fisheries Research Institute to the Tokyo Institute of Technology in accordance with international regulations under the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES). In the present study, we used a juvenile coelacanth individual TCC041-1 to determine the HOX cluster sequence. The juvenile was found in the body of female coelacanth individual (TCC041) captured off the Tanga region, Tanzania, in 2007. The capture date, locality, and identification number of the specimen were reported in Nikaido et al. (2011). To examine the insertion of a short interspersed nuclear element (SINE) at particular loci, we used the genomic DNA of six Tanzanian and two Comoran adult coelacanths (TCC035, TCC040, TCC041, TCC042, TCC043, TCC044, Comoro_1 and Comoro_2). The two Comoran coelacanth materials were provided by Kyoto University and Aquamarine Fukushima. Total genomic DNA was extracted from frozen or ethanol-preserved tissues using the DNeasy Tissue kit (Qiagen) and stored at 4 °C in TE buffer until use.

2.2. HOX cluster sequence of Indonesian coelacanth

We downloaded the currently available genomic sequence (~1.6 Mb) for Indonesian coelacanth (*L. menadoensis*) from GenBank (ID: 220898172, 220898186, 220898198, 220898210) at NCBI (Amemiya et al., 2010).

2.3. Genome sequencing

We prepared 150-bp insert paired-end libraries and single-end libraries and sequenced the samples using the Illumina genome analyzer II (GAII) according to the manufacturer's instructions. Two runs were performed. In total, 553.1 million 76-bp paired-end reads and 24.6 million 76-bp single-end reads of *L. chalumnae* genome fragments were collected. The accession number for the sequence data of this study is SRP011573.

2.4. Mapping and identification of repetitive sequences

All reads were aligned to the HOX cluster sequences of *L. menadoensis* using Burrows–Wheeler Alignment software (Li and Durbin, 2009). Each alignment was assigned a mapping quality score by Burrows–Wheeler Alignment, which is the Phred score that the alignment is incorrect. The PCR amplification step will lead to the sequencing of identical DNA fragments. Not removing these PCR duplicates can lead to the miscalling of single nucleotide variants (SNVs) by overrepresentation of one allele. This was corrected by a quality control step to remove these potential PCR duplicates with SAMtools software (Li et al., 2009). Known repetitive elements were defined by RepeatMasker (http://www.repeatmasker.org/), and novel repetitive elements were identified by pairwise alignment with the lastz program (Schwartz et al., 2003).

2.5. Calling of SNVs and short insertions/deletions (indels)

After all the reads were aligned to the reference genome using Burrows–Wheeler Alignment (Li and Durbin, 2009), we used the SAMtools to produce a consensus genotype for each genomic position. SNVs and indels were then identified based on the differences between the consensus genotype of *L. chalumnae* and the *L. menadoensis* allele at that position. These SNVs and indels were then filtered by the SAMtools variation filter, changing the minimum and maximum read depth parameter to call variants from its default values (3 and 100) to 6 and 39, to prevent the inclusion of SNVs and indels at repetitive regions. The lists of SNVs/indels were then annotated by custom Perl scripts with the SQLite database, which was specifically designed to annotate the identified SNVs/indels using information from GenBank.

2.6. Validation of SNVs

To validate the ambiguous SNVs with low conQ values, we performed PCR and direct sequencing. The PCR protocol consisted of 30 cycles with denaturation at 94 °C for 30 s, annealing at 55 °C for 45 s, and extension at 72 °C for 1 min. The PCR mixture contained 2.5 U Ex Taq polymerase™ (Takara), 1× Ex Taq buffer, 0.4 mM dNTPs, 0.1 μM of each primer, and 10 ng of template genomic DNAs in a final volume 20 μl. PCR products were confirmed by electrophoresis in a 3.0% agarose gel (Takara) and stained with ethidium bromide. The PCR products were then purified through precipitation with isopropanol. Purified PCR products were used for direct sequencing, with 25 cycles of denaturation at 96 °C for 30 s, annealing at 50 °C for 15 s, and extension at 60 °C for 1 min. Reactions contained 1 μl BigDye® ver. 3.1 terminator premix (Applied Biosystems), 1× sequencing buffer (Applied Biosystems), 1 μM sequence primer, and 2 μl of purified PCR product in a final volume of 5 μl. Sequences were determined using an automated sequencer (Applied Biosystems, model 3130).

2.7. Evolutionary analyses

The sequences were edited by GENETYX-Windows version 10 (GENETYX). Clustal W (Thompson et al., 1994) was used to align deduced amino acid sequences of HOX coding sequences of the two coelacanths, human, chimpanzee, mouse, and rat. For nucleotide sequence comparison, CodonAlign 2.0 (http://www.sinauer.com/hall/2e/) was used to introduce gaps into HOX coding sequences at the positions corresponding to the gaps in the aligned protein sequences. MEGA 5.0 software (Tamura et al., 2011) was used for genetic distance calculation for aligned HOX coding nucleotide sequences. The values of the synonymous divergence between mouse–rat, human–chimp and two coelacanth pairs were estimated and plotted for each HOX gene separately. Because the synonymous divergences were 0 in some comparisons due to the absence of no synonymous substitution, the ratios of dN/dS were estimated by concatenating the sequences of all HOX genes used for the analysis.

2.8. SINE insertion analysis

To confirm the insertion of SINEs in the coelacanths in the western Indian ocean (Tanzania and Comoros), we examined 16 loci, into which LF SINEs (Bejerano et al., 2006) were apparently inserted in the HOX cluster sequence of Tanzanian coelacanth. PCR primers were designed to amplify the SINE unit and its flanking genomic region using Primer3 (Rozen and Skaletsky, 2000). The primer sequences are summarized in Table S1. The PCR protocol was mostly the same as that shown in "Validation of SNVs" (2.6).

3. Results and discussion

3.1. Whole-genome sequencing and mapping

We performed two sequencing runs on the Illumina Genome Analyzer II platform and produced 85.9 Gbp of sequence data (Table 1). The short reads were aligned with the Burrows–Wheeler Alignment tool (Li and Durbin, 2009); we mapped 2.8% of the total reads to the *L. menadoensis* HOX cluster sequences, and 1.6% of the total reads were uniquely mapped. Insert size distribution showed not only major broad peaks around 150 bp but also a minor peak around 320 bp. This minor peak was created by the reads derived from repetitive sequences (Fig. 1; discussed in Section 3.3).

3.2. Repetitive elements and coverage

We also examined the distribution of repetitive elements defined by RepeatMasker, estimating a rate of repetitive element occupancy of ~13.1% (211,850 bp) in the HOX cluster sequences. Coverage of HOX clusters differed widely from region to region (Fig. 2). Especially, repetitive regions showed high coverage because all reads that were derived from repetitive sequences of the coelacanth genome were mapped to repetitive elements within that HOX cluster (Fig. 3). On the other hand, non-repetitive exonic regions showed a mean coverage of 22.2 (±5.9 s.d.) (Fig. 3A). These results indicate that the sequences of the exons of HOX cluster genes are probably unique in the coelacanth genome because the genome size estimated from the coverage of those exons was consistent with the size estimated from the weight of nuclear DNA content (~3 Gbp) (Makapedua et al., 2011).

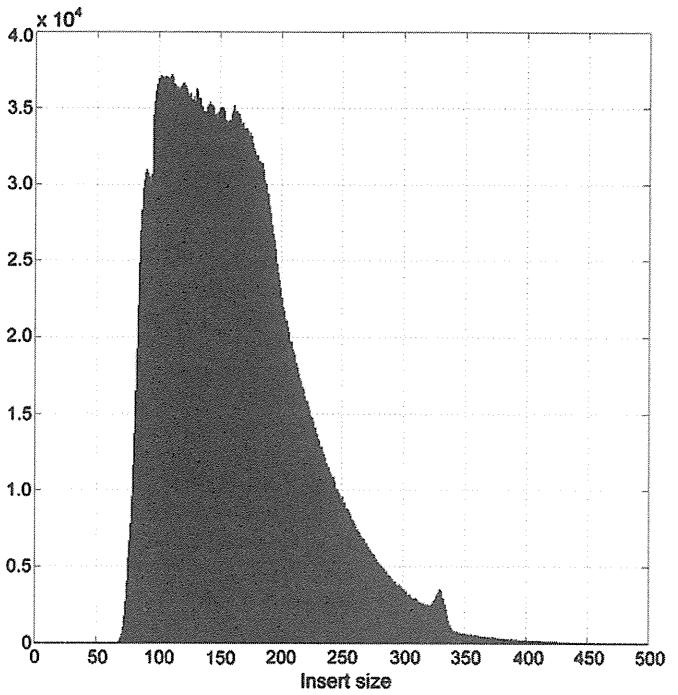

Fig. 1. Insert size distribution of paired-end reads. Insert size distribution showed two peaks, of which the minor peak at ~320 bp was derived from reads from repetitive sequences.

Table 1
Data production and mapping results for the coelacanth genome.

Data type	Number of reads	Number of mapped reads	Total bases (Mb)	Mapped bases (Mb)	Percentage with unique placement
SE	24,649,274	471,070	1873	36	84.18
PE	553,079,477	15,690,405	84,068	2385	55.93
Total	577,728,751	16,161,475	85,941	2421	56.75

Single-end (SE) and paired-end (PE) sequencing reads were aligned onto the HOX cluster sequence of *L. menadoensis* (~1.6 Mbp). 'Unique placement' means a read had only one best placement.

3.3. Identification of novel coelacanth-specific repetitive sequences

Although we excluded known repetitive elements using RepeatMasker, ~20% of intergenic sequences still had high coverage (blue line in Fig. 3C). We extracted those sequences and compared them with themselves. A dot plot provided evidence for the existence of other repetitive elements in the coelacanth genomes (Fig. 4). These elements were 2336 bp (c1) and 7736 bp (c2) in length. Although we did not characterize c1, we identified c2 to be Harbinger transposons, which were reported recently by Smith et al. (2012). As shown in Fig. 2, paired-end reads with an insert size of ~320 bp colocalized with previously identified coelacanth-specific repeats (red boxes) as well as with novel repeats identified in this study (green boxes). Their abundance suggests that these sequences are active mobile elements in the coelacanth genome. This notion is consistent with the study by Smith et al. (2012), in which they showed that the Harbinger transposons are still active *in vivo*.

3.4. Identifying SNVs and small indels

Assuming that sequences of the exons of *HOX* genes are unique (non-repetitive) in the coelacanth genome, we can define the unique region using the coverage of these exons as a reference. According to the read depth of the exons, we defined the regions with a mean exon depth ±3 s.d. as unique. After excluding extremely low and high read coverage regions, genetic differences between our sequenced *L. chalumnae* genome and the *L. menadoensis* genome were identified using modified settings in SAMtools (Li et al., 2009). We identified

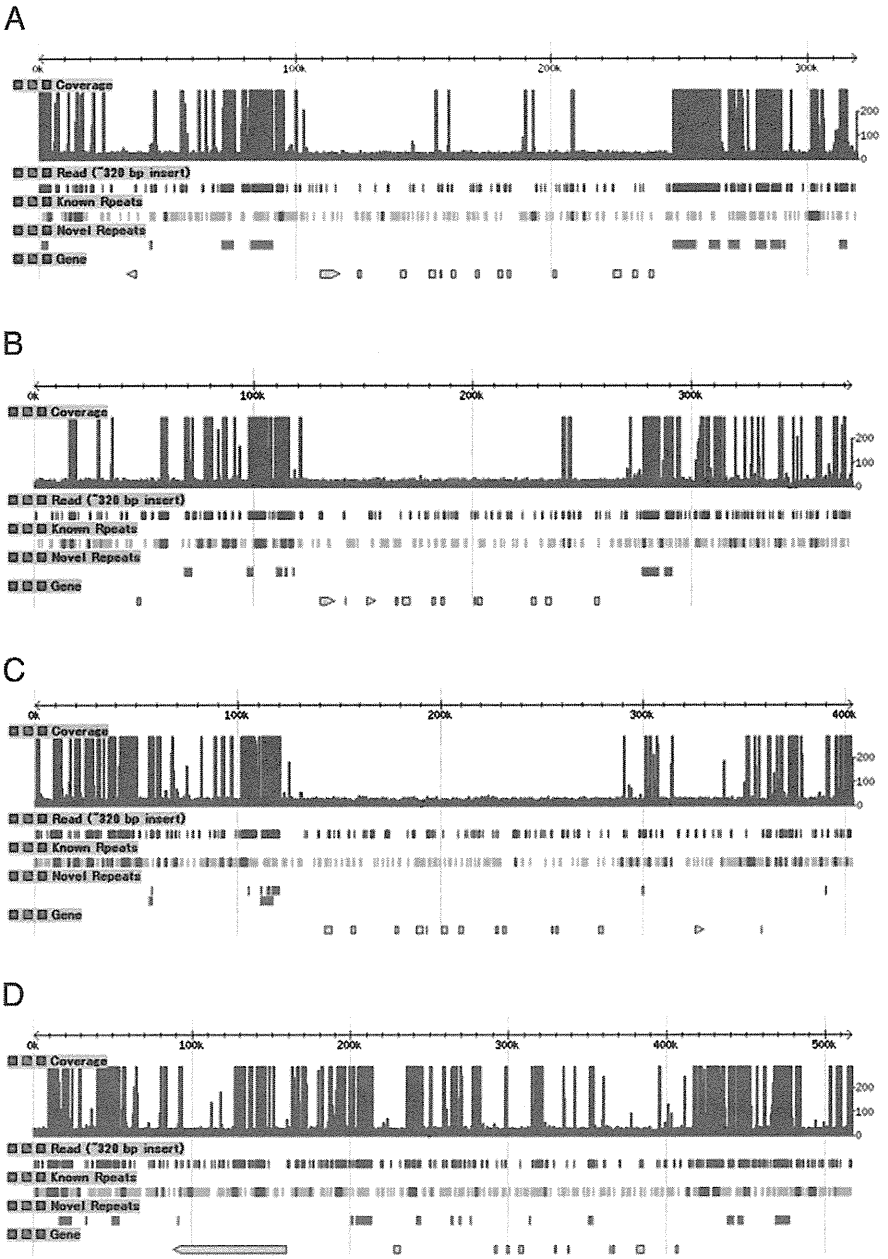

Fig. 2. Overview of HOX cluster regions. From top to bottom track, coverage of short reads, paired-end reads with ~320-bp insert size, known repetitive elements, novel repetitive elements identified in this study, and genes are shown for the HoxA (A), HoxB (B), HoxC (C), and HoxD (D) clusters. Red boxes in the known repetitive element track show previously identified coelacanth-specific repetitive sequences.

2280 SNVs and 837 indels, in which 18, 273, and 2826 variants were located in exons, introns, and intergenic regions, respectively (Table 2).

3.5. Validation of SNVs in exons

Although we filtered the possible errors during the process of variant calling, some of the SNVs shown in Table 2 were still ambiguous in terms of the low-quality scores (about 25 in conQ). To more accurately estimate the subsequent genetic divergence of HOX coding exons between the two coelacanth species, we needed to validate these SNVs to eliminate possible misidentified SNVs. Therefore, we amplified such genomic regions by PCR using the genomic DNA of the juvenile coelacanth, TCC041-1, as the template. Then, we directly determined the sequences of these PCR products by Sanger sequencing. We found that one heterozygous SNV and one homozygous indel shown in Table 2 were absent in the sequence, indicating that they were misidentified variations not eliminated during the filtering process. Accordingly, we used the modified sequences for further analyses. The modified number of SNVs and indels in coding exons is shown in parentheses (Table 2).

3.6. Estimation and comparison of synonymous divergence

Next, we estimated the genetic divergence at synonymous sites in coelacanths by focusing on whether the substitution rate is slow even at these sites. To examine this possibility, we compared the synonymous divergence of the two coelacanth species with that of the human–chimpanzee pair and that of the mouse–rat pair. Forty orthologous *HOX* genes (including two *EVX* genes) were common in all six species, whereas HOXA14, HOXB10, HOXC1, and HOXC3 were absent in mammals, and HOXD9 was absent in coelacanths.

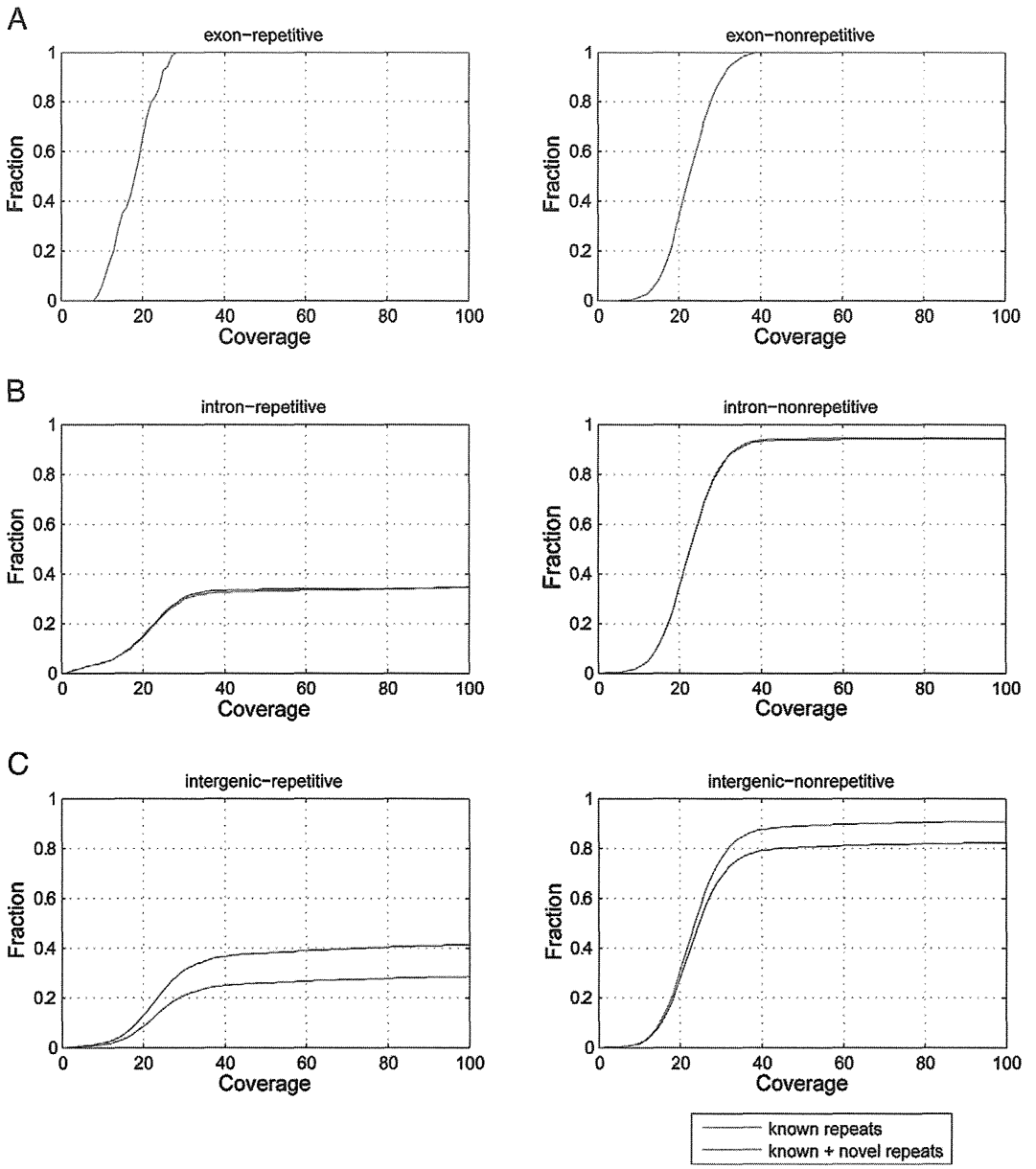

Fig. 3. Cumulative distribution of HOX cluster coverage. Cumulative distribution of read coverage in exons (A), introns (B), and intergenic (C) regions was plotted separately for repetitive (left) and non-repetitive (right) regions. Repetitive sequences were defined by RepeatMasker. Twenty percent of non-repetitive intergenic regions have extremely high coverage (blue line in C), of which 50% were derived from novel repetitive elements identified in this study (red line in C), but the remaining was unknown.

Furthermore, the putative cDNA sequences of rat HOXD8 and HOXD11 were incomplete because of the low quality of the genome sequence data for the corresponding region. Accordingly, we removed them from the analysis. In total, we used the 38 *HOX* gene orthologs for the subsequent analyses to compare the same genomic region among six species. The protein sequences were first aligned using Clustal W implemented in MEGA 5 with the default settings. Then, gaps were introduced into the cDNA sequences based on the protein sequence alignment. The aligned nucleotide sequences were used to estimate genetic divergence using MEGA 5. Fig. 5 shows the plots of the uncorrected synonymous divergence (y axis) against the corrected nucleotide divergence (x axis) of 38 *HOX* genes for each of the three species pairs (human–chimp, mouse–rat, and Tanzania–Indonesia coelacanths), indicating no signs of saturation of synonymous substitutions. Accordingly, we could compare the synonymous divergence among the three species pairs without regard to the possibility of saturation.

Table 3 shows the resultant averaged estimated synonymous divergence of *HOX* genes for each of the three species pairs. In the human–chimp and mouse–rat pairs, the synonymous divergence was 0.78% and 7.5%, respectively. Although these values are smaller than those estimated at the whole-genome level (human–chimp 1.1%, Chen et al., 2001; mouse–rat 19%, Rat Genome Sequencing Project Consortium, 2004), they are consistent with previous studies in which the authors indicated that *HOX* genes were conserved even at synonymous sites (Lin et al., 2008; Woltering and Duboule, 2009). The most intriguing result is that the synonymous divergence was much smaller in the coelacanth species pair; the value was 0.07%, which is about 11-fold smaller than that of the human–chimp pair.

When we consider that the divergence time of human–chimp and the two coelacanth species is almost the same (see Section 3.7), the difference in the synonymous divergence between these groups was statistically significant ($p<10^{-14}$ by two-sample proportion z-test, Snedecor and Cochran, 1989).

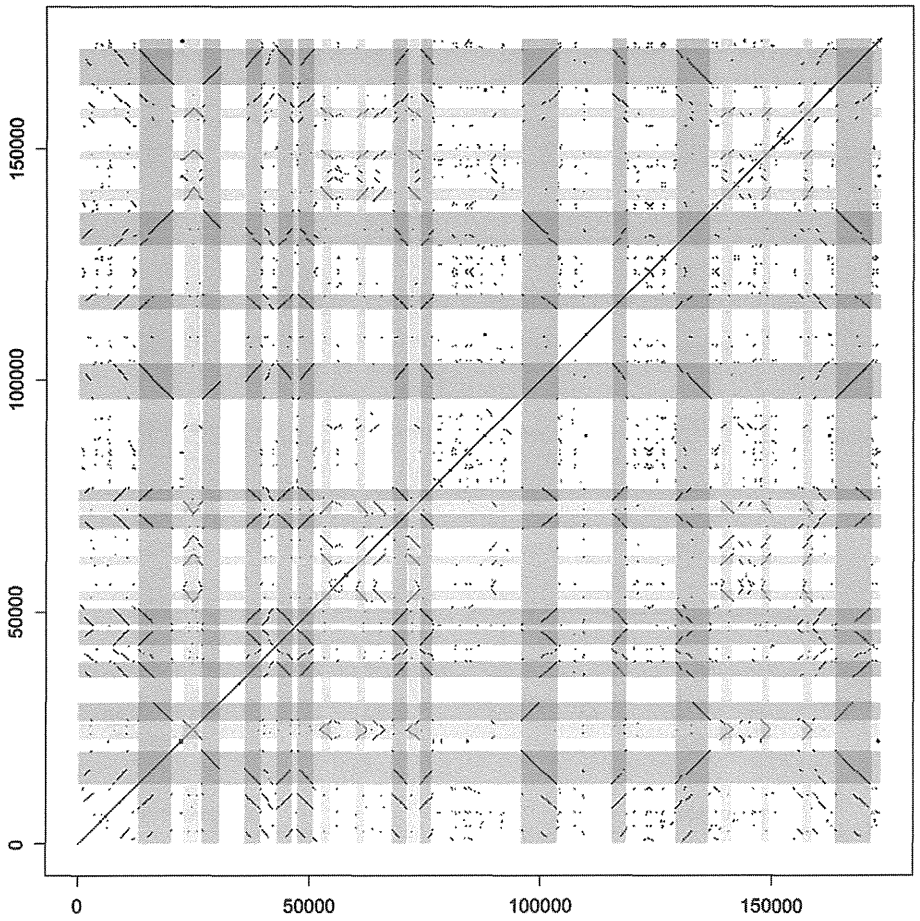

Fig. 4. Dot plot of HOX cluster region. After excluding known repetitive sequences identified by RepeatMasker, the concatenated remaining sequence was compared with itself by lastz (Schwartz et al., 2003). Green and red shadows indicate two newly identified repetitive sequences specific to coelacanth.

3.7. Estimation and comparison of synonymous substitution rates

To directly compare the absolute values of the synonymous substitution rate among the three species pairs, we interpolated the divergence times estimated by the independent studies. We used 6 MYA and 16 MYA for the divergence of human–chimp (Glazko and Nei 2003) and mouse–rat (Hasegawa et al., 2003) pairs, respectively. For coelacanths, there are two main alternative hypotheses for their divergence time (6 MYA and 30 MYA), which have not been validated (Holder et al., 1999; Inoue et al., 2005; Sudarto et al., 2010). The estimated value of 6 MYA by Holder et al. (1999) appears to be less reliable than that of 30 MYA by Inoue et al. (2005) because the former study analyzed just partial mitochondrial sequences and applied the substitution rate of tetrapods. However, we here used both 6 MYA and 30 MYA to see the difference between the resultant values. The synonymous substitution rates of human–chimp and mouse–rat were calculated to be 0.65×10^{-9} and 2.33×10^{-9} (per site per year), respectively (Table 3). The higher substitution rate in rodents relative to that of primates is consistent with previous studies (Kitano and Saitou, 2005; Li et al., 1987). As for the coelacanths, when we applied the divergence time of 6 MYA, the substitution rate was calibrated to be 0.059×10^{-9} (per site per year, Table 3), which is 11-fold slower than that of human–chimp. Furthermore, the divergence time of 30 MYA, which appears to be more reliable dating, led to the significantly slow substitution rate of 0.0012×10^{-9} (per site per year, Table 3). This value is 56-fold slower than that of human–chimp.

Table 2
SNVs and indels in unique (non-repetitive) regions.

Locus	SNV		Indel		Total	Percentage	Length of region (bp)
	Homo	Hetero	Homo	Hetero			
Exon	15	1 (0)	2 (1)	0	18 (16)	0.0445% (0.0395%)	40,471
Intron	147	15	104	7	273	0.2373%	115,045
Intergenic	1444	658	695	29	2826	0.2514%	1,124,218
Total	1606	674	801	36	3117	0.2515%	1,239,263

The modified number of SNVs and indels in coding exons after validation is shown in parentheses (see Section 3.5).

3.8. SINE insertion polymorphisms

To test whether the insertion of transposable elements occurred after the split of *L. chalumnae* and *L. menadoensis*, we searched the alignments of the four HOX clusters for indels between these two species. Six indels of >50 nt were identified, in which the sequences were found in *L. menadoensis* but were absent in *L. chalumnae*. However, these differences did not result from insertions of transposable elements (data not shown). Specifically, three of the six differences were copy number variations of tandem duplications of a ~60-nt unit, and two of them were due to partial deletions of Harbinger-like transposons in *L. chalumnae*. The remaining sequence showed no homology with known transposable elements. This result suggests that there is no difference in the presence/absence pattern of transposable elements in the HOX clusters. However, owing to the mapping strategy of short read fragments, it is still possible that the polymorphisms of the short-length retroposons such as SINEs were actually absent. To examine such a possibility, we examined 16 loci, into which LF SINEs (Bejerano et al., 2006) were apparently inserted in *L. chalumnae*.

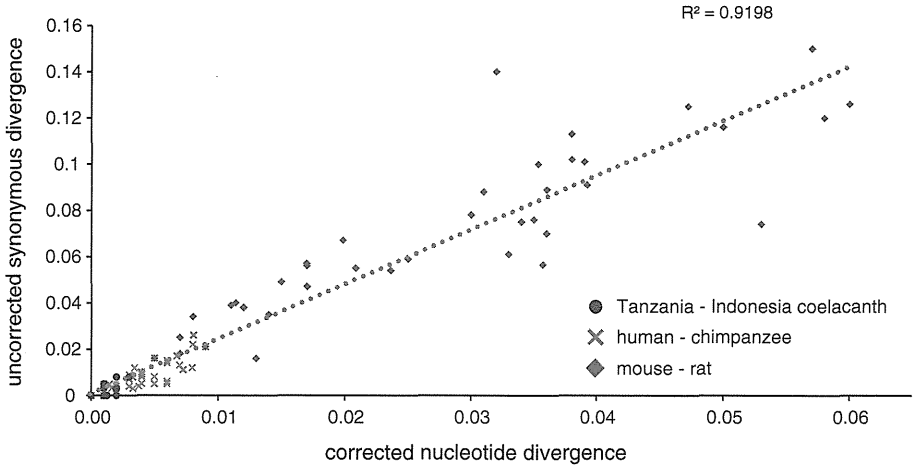

Fig. 5. Saturation plot for the synonymous divergence of *HOX* genes of three species pairs. The uncorrected synonymous divergence from the pairwise comparison plotted against the pairwise corrected nucleotide divergence. The synonymous divergence increases almost linearly, indicating that the saturation is apparently absent in this comparison. The uncorrected synonymous divergence was estimated using the modified Nei–Gojobori method, and the corrected nucleotide divergence was estimated with the maximum composite likelihood method.

The 16 SINEs were chosen to represent recently inserted copies based on the divergence from the consensus sequence. Then, we performed SINE-flanking PCR to investigate the presence or absence of a SINE at each locus. If a SINE is absent at a particular locus, the PCR band becomes shorter than expected (Okada et al., 2004). We detected SINE insertions in all of the loci examined (Fig. 6), indicating very few, if any, SINE insertion polymorphisms between *L. chalumnae* and *L. menadoensis*. Accordingly, low genetic diversity of the HOX cluster region between the two coelacanth species was further indicated in terms of SINE insertions.

3.9. Reduction of substitution rate, purifying selection, or recent divergence?

An important question is why the genetic divergence of the coelacanth HOX cluster is so small, even at the synonymous sites. There are three possible explanations: (1) strong purifying selection in the open reading frames of *HOX* genes, including the synonymous sites, (2) reduction of the nucleotide substitution rate in the coelacanth lineage, and (3) exceptionally more recent divergence between *L. chalumnae* and *L. menadoensis* than expected from the mtDNA analyses. Regarding the first explanation, Lin et al. (2008) indicated that the mammalian *HOX* genes have ultraconserved coding regions, in which the sequences are highly conserved at the nucleotide level. Their study implied that the ultraconserved coding regions have some gene regulatory functions, which led to purifying selection without regard to synonymous and non-synonymous sites. Accordingly, it is possible that the smaller genetic divergence in coelacanth *HOX* genes may reflect substantially stronger purifying selection compared with that in rodents and primates. However, the dN/dS ratios of *HOX* genes among human–chimp, rodents and two coelacanth pairs were 0.23, 0.42 and 0.39, respectively, indicating that the purifying selection in coelacanth was not strong in comparison with the other groups. Accordingly, the smaller synonymous divergence in the coelacanth *HOX* genes may not be caused by the operation of stronger purifying selection. Regarding the second possibility, an exceptionally slow substitution rate has been reported in the mtDNA of shark, which is 7- to 8-fold slower than that of primates or ungulates (Martin et al., 1992). Given the similarity of coelacanths to sharks in their ecology, large body size, and long generation time, it is plausible that coelacanths also exhibit a slow nucleotide substitution rate in their genomes similar to that observed in sharks. The small genetic divergence even at introns and intergenic regions (0.25%, Table 2) of coelacanth *Hox* genes and the apparent absence of SINE insertion polymorphisms, are consistent with this possibility. However, there are also arguments for the third possibility, namely, the more recent divergence time of *L. chalumnae* and *L. menadoensis*. Because coelacanths have been a single lineage since the Devonian era, they do not have closely related taxa, making it difficult to fix the calibration point for divergence time estimation (Sudarto et al., 2010). Although there is a large difference in the estimates between the two major hypotheses, even the later estimate of 6 MYA leads to an 11-fold slower substitution rate in coelacanths. However, if the two coelacanth species diverged much more recently than what has been estimated, the conclusion would be drastically changed. Indeed, Schartl et al. (2005) raised the possibility that the western Indian Ocean could have been colonized recently by coelacanth drifters from the Pacific Province by Indonesian throughflow, which has probably existed for only 3 to 4 MY. Gordon (1998) also pointed out the presence of an oceanographic connection between Sulawesi and the Comoran Islands region that has enabled very recent gene flow from Indonesia to Comoros. Furthermore, if such dispersal occurred only in male individuals, the genetic divergence should be smaller in the nuclear genome than in the mitochondrial genome. This may provide us with a compromise resolution for the exceptionally smaller genetic divergence in the HOX cluster of coelacanths compared with the divergence time based on mtDNA.

Table 3
Estimated genetic divergences and substitution rates at synonymous sites.

	Divergence time from literature	Estimated synonymous divergence	Estimated synonymous substitution rate (per site per year)
Human–chimp	6 MYA[a]	0.0078 (0.0012)[b]	$0.65\ (0.10)\times10^{-9}$
Tanzanian–Indonesian coelacanths	6 MYA[c]	0.00070 (0.00011)	$0.059\ (0.0090)\times10^{-9}$
	30 MYA[d]		$0.012\ (0.0018)\times10^{-9}$
Mouse–rat	16 MYA[e]	0.075 (0.012)	$2.33\ (0.37)\times10^{-9}$

[a] Inoue et al. (2005).
[b] The values in parentheses indicate the standard error.
[c] Holder et al. (1999).
[d] Glazko and Nei (2003) and Sudarto et al. (2010).
[e] Hasegawa et al. (2003).

3.10. Conclusion

In the present study, we performed a large-scale comparison of the HOX cluster sequence between two coelacanth species, showing a significantly small genetic divergence, even at silent sites, relative

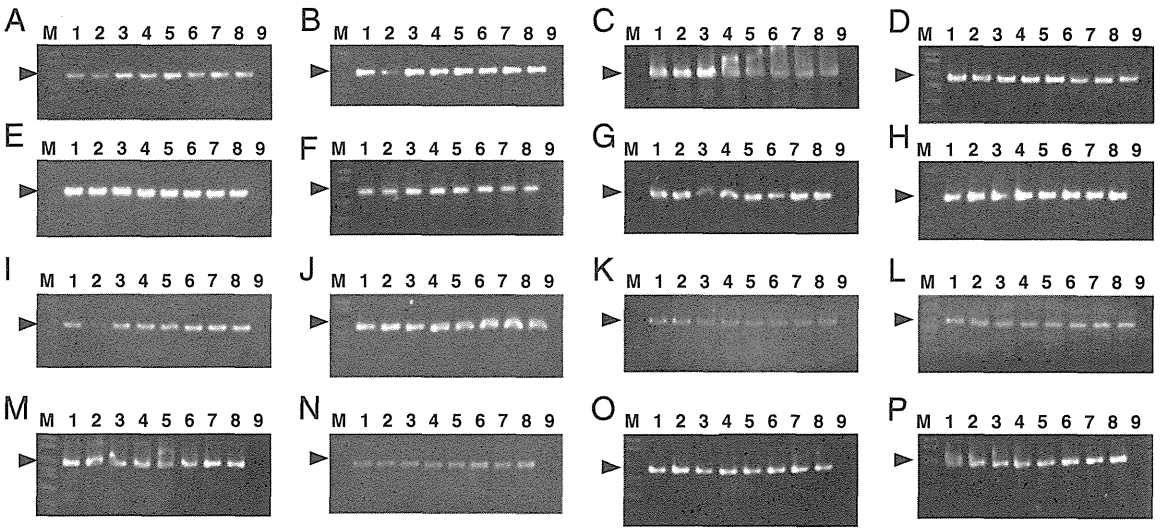

Fig. 6. SINE insertion analysis for the 16 loci of coelacanths in the western Indian Ocean. (A–P) Agarose gel electrophoresis of SINE-flanking PCR products from locus 1 to 16. Lane M indicates the size marker, and lanes 1–9 indicate the *L. chalumnae* samples (ID: Comoro_1, Comoro_2, TCC039, TCC040, TCC041, TCC042, TCC043, TCC044, and negative control). The arrowheads indicate the expected size of the PCR products, which were deduced from the sequence of *L. chalumnae*. In all loci, PCR bands were detected at ~600 bp, which is expected for the presence of a SINE.

to human–chimp and mouse–rat pairs. When we applied the divergence times deduced by the mtDNA sequence data, the substitution rate was calibrated to be significantly slower than that of the other vertebrates reported. To explain the phenomenon, we propose two alternative possibilities, which are not necessarily mutually exclusive: reduction of the nucleotide substitution rate in coelacanth lineages or an unexpectedly recent divergence of the two coelacanth species. Although it is difficult to settle this issue at present, whole genome-wide analysis and large-scale population genetic analyses, which will become available soon, may shed light on the subject. Such genome-wide analyses may eventually elucidate why coelacanths could be evolutionary relics from the Devonian time.

Acknowledgments

This work was supported by the JSPS AA Science Platform Program, a Grant-in-Aid for Scientific Research (S) from the Ministry of Education, Culture, Sports, Science and Technology of Japan (to N.O.), and the Global COE Program "Deciphering Biosphere from Genome Big Bang" (to S.M.).

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.gene.2012.05.047.

References

Amemiya, C.T., Powers, T.P., Prohaska, S.J., Grimwood, J., Schmutz, J., Dickson, M., Miyake, T., Schoenborn, M.A., Myers, R.M., Ruddle, F.H., Stadler, P.F., 2010. Complete HOX cluster characterization of the coelacanth provides further evidence for slow evolution of its genome. Proc. Natl. Acad. Sci. U. S. A. 107, 3622–3627.

Bejerano, G., Lowe, C.B., Ahituv, N., King, B., Siepel, A., Salama, S.R., Rubin, E.M., Kent, W.J., Haussler, D., 2006. A distal enhancer and an ultraconserved exon are derived from a novel retroposon. Nature 441, 87–90.

Carroll, R.L., 1988. Vertebrate Paleontology and Evolution. H. Freeman and Co., New York.

Chen, F.C., Vallender, E.J., Wang, H., Tzeng, C.S., Li, W.H., 2001. Genomic divergence between human and chimpanzee estimated from large-scale alignments of genomic sequences. J. Hered. 92, 481–489.

De Vos, L., Oyugi, D., 2002. First capture of a coelacanth, *Latimeria chalumnae* Smith, 1939 (Pisces: Latimeriidae), off Kenya. S. Afr. J. Sci. 98, 345–347.

Erdman, M.V., Caldwell, R.L., Moosa, M.K., 1998. Indonesian 'king of the sea' discovered. Nature 395, 335.

Glazko, G.V., Nei, M., 2003. Estimation of divergence times for major lineages of primate species. Mol. Biol. Evol. 20, 424–434.

Gordon, A.L., 1998. Coelacanth populations may go with the flow. Nature 395, 634.

Hasegawa, M., Thorne, J.L., Kishino, H., 2003. Time scale of eutherian evolution estimated without assuming a constant rate of molecular evolution. Genes Genet. Syst. 78, 267–283.

Heemstra, P.C., Freeman, A.L., Wong, H.Y., Hensley, D.A., Rabesandratana, H.D., 1996. First authentic capture of a coelacanth, *Latimeria chalumnae* (Pisces: Latimeriidae), off Madagascar. S. Afr. J. Sci. 92, 150–151.

Holder, M.T., Erdmann, M.V., Wilcox, T.P., Caldwell, R.L., Hillis, D.M., 1999. Two living species of coelacanths? Proc. Natl. Acad. Sci. U. S. A. 96, 12616–12620.

Inoue, J.G., Miya, M., Venkatesh, B., Nishida, M., 2005. The mitochondrial genome of Indonesian coelacanth *Latimeria menadoensis* (Sarcopterygii: Coelacanthiformes) and divergence time estimation between the two coelacanths. Gene 349, 227–235.

Kitano, T., Saitou, N., 2005. Evolutionary conservation of 5′ upstream sequence of nine genes between human and great apes. Genes Genet. Syst. 80, 225–232.

Krumlauf, R., 1994. *Hox* genes in vertebrate development. Cell 78, 191–201.

Kurosawa, G., Takamatsu, N., Takahashi, M., Sumitomo, M., Sanaka, E., Yamada, K., Nishii, K., Matsuda, M., Asakawa, S., Ishiguro, H., Miura, K., Kurosawa, Y., Shimizu, N., Kohara, Y., Hori, H., 2006. Organization and structure of hox gene loci in medaka genome and comparison with those of pufferfish and zebrafish genomes. Gene 370, 75–82.

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25, 1754–1760.

Li, W.H., Tanimura, M., Sharp, P.M., 1987. An evaluation of the molecular clock hypothesis using mammalian DNA sequences. J. Mol. Evol. 25, 330–342.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079.

Lin, Z., Ma, H., Nei, M., 2008. Ultraconserved coding regions outside the homeobox of mammalian Hox genes. BMC Evol. Biol. 8, 260.

Maisey, J.G., 1996. Discovering Fossil Fishes. Henry Holt, New York.

Makapedua, D.M., Barucca, M., Forconi, M., Antonucci, N., Bizzaro, D., Amici, A., Carradori, M.R., Olmo, E., Canapa, A., 2011. Genome size, GC percentage and 5mC level in the Indonesian coelacanth *Latimeria menadoensis*. Mar. Genomics 4, 167–172.

Martin, A.P., Naylor, G.J., Palumbi, S.R., 1992. Rates of mitochondrial DNA evolution in sharks are slow compared with mammals. Nature 357, 153–155.

Meyer, A., Málaga-Trillo, E., 1999. Vertebrate genomics: more fishy tales about *Hox* genes. Curr. Biol. 9, R210–R213.

Nikaido, M., Sasaki, T., Emerson, J.J., Aibara, M., Mzighani, S.I., Budeba, Y.L., Ngatunga, B.P., Iwata, M., Abe, Y., Li, W.H., Okada, N., 2011. Genetically distinct coelacanth population off the northern Tanzanian coast. Proc. Natl. Acad. Sci. U. S. A. 108, 18009–18013.

Noonan, J.P., Grimwood, J., Danke, J., Schmutz, J., Dickson, M., Amemiya, C.T., Myers, R.M., 2004. Coelacanth genome sequence reveals the evolutionary history of vertebrate genes. Genome Res. 14, 2397–2405.

Okada, N., Shedlock, A.M., Nikaido, M., 2004. Retroposon mapping in molecular systematics. Methods Mol. Biol. 260, 189–226.

Rat Genome Sequencing Project Consortium, 2004. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. Nature 428, 493–521.

Rozen, S., Skaletsky, H.J., 2000. Primer3 on the WWW for general users and for biologist programmers. In: Krawetz, S., Misener, S. (Eds.), Bioinformatics Methods and Protocols: Methods in Molecular Biology. Humana Press, Totowa, NJ, pp. 365–386.

Schartl, M., Hornung, U., Hissmann, K., Schauer, J., Fricke, H., 2005. Genetics: relatedness among east African coelacanths. Nature 435, 901.

Schliewen, U., Fricke, H., Schartl, M., Epplen, J.T., Pääbo, S., 1993. Which home for coelacanth? Nature 363, 405.

Schwartz, S., Kent, W.J., Smit, A., Zhang, Z., Baertsch, R., Hardison, R.C., Haussler, D., Miller, W., 2003. Human–mouse alignments with BLASTZ. Genome Res. 13, 103–107.
Smith, J.L.B., 1939. A living fish of Mesozonic type. Nature 143, 455–456.
Smith, J.L.B., 1953. The second Coelacanth. Nature 171, 99–101.
Smith, J.J., Sumiyama, K., Amemiya, C.T., 2012. A living fossil in the genome of a living fossil: harbinger transposons in the coelacanth genome. Mol. Biol. Evol. 29, 985–993.
Snedecor, G.W., Cochran, W.G., 1989. Statistical Methods, 8 edn. Iowa State University Press.
Sudarto, Lalu, X.C., Kosen, J.D., Tjakrawidjaja, A.H., Kusumah, R.V., Sadhotomo, B., Kadarusman, Pouyaud, L., Slembrouck, J., Paradis, E., 2010. Mitochondrial genomic divergence in coelacanths (*Latimeria*): slow rate of evolution or recent speciation? Mar. Biol. 157, 2253–2262.
Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28, 2731–2739.
Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680.
Woltering, J.M., Duboule, D., 2009. Conserved elements within open reading frames of mammalian Hox genes. J. Biol. 8, 17.