Mutation Patterns of Human Proto-Oncogene Tyrosine-Protein Kinase Receptor RET by Means of Amino-Acid Pair Predictability

Qingwu Yu[1], Shaomin Yan[2], Guang Wu[2*]

[1]Department of Business Administration, Guangxi Economic Management Cadre College, 105 University Avenue, Nanning, Guangxi 530007, China
[2]State Key Laboratory of Non-food Biomass Enzyme Technology, National Engineering Research Center for Non-food Biorefinery, Guangxi Key Laboratory of Biorefinery, Guangxi Academy of Sciences, 98 Daling Road, Nanning, Guangxi 530007, China
*Email: hongguanglishibahao@yahoo.com

Abstract: Mutation is the driving force for protein evolution, however, it is still unlikely to use a very simple way to generalize the mutation patterns. In this study, we used the amino-acid pair predictability to classify the human RET protein with its 130 single missense variants, and then to determine which amino-acid pair was more sensitive to mutations. The results showed that (i) the unpredictable amino-acid pairs are more sensitive to mutations, (ii) most mutations target the original amino-acid pairs whose actual frequency is larger than their predicted one, (iii) most mutations form the substituting amino-acid pairs whose actual frequency is smaller than their predicted one, and (iv) mutations generally narrow the difference between actual and predicted frequencies of affected amino-acid pairs, which are consistent with our previous studies, providing a very simple way to generalize mutations patterns and the underlined mechanism driving mutations, and highlighting that the main cause for protein mutation is highly likely to be the force to narrow the difference between predicted and actual frequencies of amino-acid pairs.
[Qingwu Yu, Shaomin Yan, Guang Wu. **Mutation Patterns of Human Proto-Oncogene Tyrosine-Protein Kinase Receptor RET by Means of Amino-Acid Pair Predictability**. *Life Sci J* 2013; 10(3): 1667-1672]. (ISSN: 1097-8135). http://www.lifesciencesite.com 251

Keywords: Amino-acid pair predictability; human proto-oncogene tyrosine-protein kinase receptor RET; mutation pattern

1. Introduction

Mutation is the driving force behind many processes linked to human diseases, including cancer [1-3], aging [4-6], and the evolution of drug resistance [7]. Great efforts have been made to illustrate the underlined mechanisms for these mutations [8-14], however, it is still unlikely to use a very simple way to generalize the mutation patterns because the underlined mechanisms seem to be different case by case. In more plain words, we literally cannot use a single sentence to answer the question of why mutations occur.

This argument is important because the process for humans to understand the nature follows the circle from the simple to the complex, and then from the complex to the simple that is abstracted from the complex. We must stop somewhere, and for science to be possible we must stop where we have found simplicity [15], which could be applicable to the mechanism of mutation. Over the last decade, computational mutation approach has been developed according to random mechanism, which include three methods to compute the chance of mutation: amino-acid pair predictability, amino-acid distribution probability and amino-acid mutating probability (for the details, see the review articles [16-18] and the monograph [19]).

We have used the amino-acid pair predictability to analyze mutations from various human proteins including, adrenoleukodystrophy protein [20], androgen receptor [21], Bruton's tyrosine kinase [22], coagulation factor IX [23], collagens [24-26], copper-transporting ATPase 2 [27], α-galactosidase A [28], β-glucocerebrosidase [29], haemoglobins [30, 31], low-density lipoprotein receptor [32], menin [33], ornithine transcarbamylase [34], p53 protein [35], phenylalanine hydroxylase protein [36], protein C [37], von Hippel-Lindau disease tumour suppressor [38]. The results using this approach showed that the mutations were characterized by four features: (i) the vast majority of mutations occurred at unpredictable amino-acid pairs, (ii) most mutations targeted the original amino-acid pairs that had actual frequency larger than predicted one, (iii) most mutations generated one or both substituting amino-acid pairs that were absent in normal protein, and (iv) most mutations narrowed the difference between actual and predicted frequencies in amino-acid pairs. These features highlight that mutations generally lead their variants to be constructed more randomly and stably. Anyway, more studies are needed to get a more sold conclusion.

The RET proto-oncogene encodes the RET

receptor tyrosine kinase, which has essential roles in cell survival, differentiation and proliferation [39, 40]. Both germline and somatic mutations of the RET proto-oncogene can induce neuroendocrine tumour, such as medullary thyroid carcinoma [41-45], multiple endocrine neoplasia type 2 [46, 47], breast cancers [48], Hirschsprung's disease [49] and so on. Efforts have been made to manufacture small-molecule tyrosine kinase inhibitors that can target multiple kinases at nanomolar concentrations, which have shown efficacy against a variety of malignancies [50-52]. To date, 137 natural variants have been found in the RET protein documented in the UniProt Knowledgebase (UniProtKB) [53], which provide another opportunity to analyze the mutation pattern using computational mutation approach.

2. Material and Methods

The amino-acid sequence of human RET protein (accession number P07949) with its 130 single missense variants is obtained from UniProtKB [53].

As being done in our previous studies [20-38], the amino-acid pair predictability was used to classify the amino-acid pairs in RET protein as predictable and unpredictable in following way. For example, the RET protein had 77 alanines (A) and 75 arginines (R) of its 1114 amino acids: according to the permutation, the amino-acid pair AR should appear 5 times ($77/1114 \times 75/1113 \times 1113 = 5.184$); indeed there were five ARs in it, so the pair AR is predictable. For another example, there are 35 glutamines (Q) in the protein: the amino-acid pair AQ should appear twice ($77/1114 \times 35/1113 \times 1113 = 2.419$) according to the permutation; actually the pair AQ appeared five times and belonged to unpredictable. In this way, all amino-acid pairs in RET protein were classified as predictable or unpredictable.

When a mutation occurred in RET protein, it was easily determined that the affected amino-acid pairs belonged to predictable or unpredictable. With a large number of documented mutations in database, the mutation patterns could be figured out in regard whether mutations occurred in either predictable amino-acid pairs or unpredictable ones statistically differently.

The *Chi*-square test was used to compare the number of amino-acid pairs, and the Mann-Whitney *U*-test was used to compare the difference between predicted and actual frequencies.

3. Results and Discussions

The human RET protein is actually composed of 1114 amino acids, which can construct 1113 adjacent amino-acid pairs, and our computation is strictly related to these 1113 adjacent amino-acid pairs. As there are only 400 types of amino-acid pairs resulting from 20 kinds of amino acids, so there must be some types of amino-acid pairs appeared several times. Consequently, the 400 types of amino-acid pairs and 1113 pairs in human RET protein are classified as predictable or unpredictable. Of 335 types of amino-acid pairs found in the protein, 94 types are predictable and 241 types are unpredictable. Of 1113 amino-acid pairs, 246 and 867 are predictable and unpredictable. Thus, the mutations were categorized into predictable type/pair and unpredictable type/pair listed in Table 1.

As Table 1 clearly demonstrates that mutations are highly likely to occur in unpredictable types as well as pairs, one might readily ask why. To answer this, we might at first look at what effect these mutations brought about. For this purpose, we would like to compare the difference between predicted frequency (PF) and actual frequency (AF) of amino-acid pairs, where the mutations occurred. For instance, the mutation at position 251 substituted glutamic Acid (E) for lysine (K), which resulted in two amino-acid pairs “RE” and “EE” mutating to “RK” and “KE”, because the amino acids were “R” and “E” at positions 250 and 252. The predicted frequency and actual one were 5 and 7 for “RE”, 5 and 10 for “EE”, 3 and 4 for “RK”, and 3 and 2 for “KE” before and after mutation. Thus the difference between predicted and actual frequencies was -7 for the pairs before mutation, i.e. $(5-7)+(5-10)$, and 0 to the pairs after mutation, i.e., $(3-4)+(3-2)$. By this way, Figure 1 demonstrated the effect of all mutations on both original and substituting amino-acid pairs.

Figure 1 suggests what the mutations did is to narrow the difference between predicted and actual frequencies. This figure actually answers the question why mutations are highly likely to occur at unpredictable types as well as pairs, because it is only the unpredictable type/pair having the difference between predicted and actual frequencies. In other words, the mutations specifically target the unpredictable types/pairs in order to narrow the difference between their predicted and actual frequencies.

To furthermore verify this rationale, it is necessary to more closely look at the original amino-acid pairs targeted by mutations and the substituting amino-acid pairs formed after mutations in Table 2, which can be read as follows. The first column classified the amino-acid pairs as original and substituting. The second and third columns showed the types of these two amino-acid pairs affected by

Table 1. Mutation occurrence in predictable and unpredictable types/pairs of human RET protein

RET protein	Type		Pair		Mutations		Ratio	
	number	%	number	%	number	%	mutations/types	mutations/pairs
Predictable	93	27.84	239	21.47	8	6.15	8/93 = 0.09	8/239 = 0.03
Unpredictable	241	72.16	874	78.53	122	93.85	122/241 = 0.51	122/874 = 0.14
Total	334	100	1113	100	130	100	130/334 = 0.39	130/1113 = 0.12

Table 2. Classification of original and substituting amino-acid pairs before and after mutation in human RET protein

Amino-acid pairs			Before Mutation			After Mutation		
	pair I	pair II	number	%	total %	number	%	total %
Original	AF = PF	AF = PF	8	6.15	6.15	14	10.77	10.77
	AF > PF	AF > PF	53	40.77	93.85	24	18.46	89.23
	AF > PF	AF = PF	21	16.15		16	12.31	
	AF > PF	AF < PF	32	24.62		38	29.23	
	AF < PF	AF = PF	15	11.54		14	10.77	
	AF < PF	AF < PF	1	0.77		24	18.46	
Substituting	AF = PF	AF = PF	8	6.15	6.15	11	8.46	8.46
	AF > PF	AF > PF	13	10.00	93.85	38	29.23	91.54
	AF > PF	AF = PF	17	13.08		43	33.08	
	AF > PF	AF < PF	43	33.08		26	20.00	
	AF < PF	AF = PF	23	17.69		10	7.69	
	AF < PF	AF < PF	26	20.00		2	1.54	

AF: actual frequency; PF: predicted frequency. There is a remarkably statistical difference before and after mutation, as well as between original and substituting pairs ($P \leq 0.001$, the *Chi*-square test).

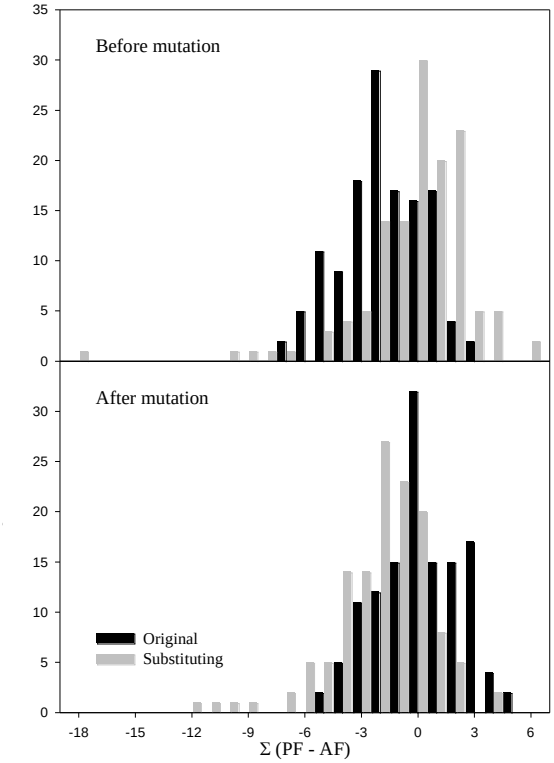

Figure1. Distribution of difference between predicted and actual frequencies in original and substituting amino-acid pairs before and after mutation in human RET protein.

PF, predicted frequency; AF, actual frequency.

mutation in relation of their predicted and actual frequencies, for example, the third and seventh rows indicated that the actual frequencies were equal to the predicated ones in both pairs I and II. The fourth and fifth columns indicated how many mutations occurred in amino-acid pairs I and II: for instance, 8 of 130 mutations (6.15%) occurred at the original amino-acid pairs whose actual frequencies was equal to their predicted ones. The sixth column indicated the percentage of mutations occurring at predictable and unpredictable amino-acid pairs. The data before mutation were listed in columns 4-6 while the corresponding values after mutation in columns 7-9.

Table 2 evidently demonstrates that mutations mainly target the original amino-acid pairs whose actual frequency is larger than their predicted one whereas they mainly form the substituting amino-acid pairs whose actual frequency is smaller than their predicted one. Thus, mutations play their role to narrow the difference between actual and predicted frequencies by means of reducing the actual frequency in original amino-acid pairs and increasing the actual frequency in substituting pairs.

Figure 2 displays the statistical results from Figure 1, which points out the changes of amino-acid pairs in the difference between predicted and actual frequencies before and after mutation. Before mutation, the median of difference between predicted and actual frequencies is –2 in original amino-acid pairs, which demonstrates that the mutations targeted

the amino-acid pairs that appeared more than their predicted frequency. Meanwhile, the corresponding value is 1 in substituting amino-acid pairs, which suggests that the mutations generated the amino-acid pairs that appeared less than their predicted frequency. Interestingly, the features of amino-acid pairs revealed opposite changes after mutation. Remark statistical difference was found between the corresponding groups ($P < 0.001$).

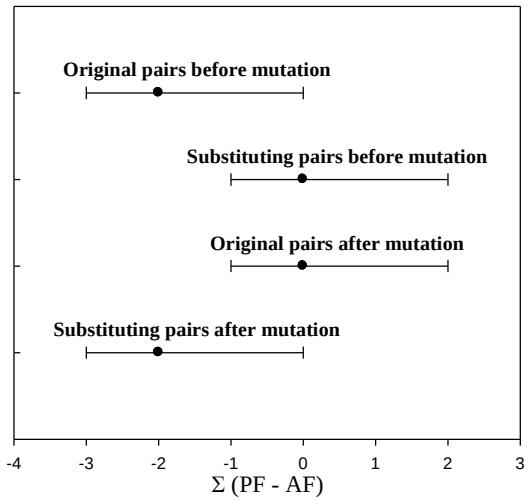

Figure2. Statistical comparison of difference between actual and predicted frequencies of original and substituting amino-acid pairs before and after mutation in human RET protein. The data are presented by median with an interquatile interval. There is statistically significant difference between corresponding groups before and after mutation ($p < 0.001$, the Mann-Whitney *U*-test).

Both internal and external factors can influence protein evolution [54, 55]. This study has addressed three issues: (i) at which amino-acid pairs mutations are more likely to occur, (ii) why mutations target these amino-acid pairs, and (iii) by which means mutations form the substituting pairs. The current results obtained from RET protein are completely consistent with our previous ones [20-38], demonstrating that (i) the unpredictable amino-acid pairs are more sensitive to mutations, (ii) most mutations target the original amino-acid pairs whose actual frequency is larger than their predicted one, (iii) most mutations form the substituting amino-acid pairs whose actual frequency is smaller than their predicted one, and (iv) mutations generally narrow the difference between actual and predicted frequencies of affected amino-acid pairs. Therefore, we can use a very simple way to generalize mutations patterns and the underlined mechanism driving mutations, that is, the main cause for protein mutation is highly likely the force to narrow the difference between predicted and actual frequencies of amino-acid pairs. Indeed, mutations eliminate old unpredictable amino-acid pairs while produce new unpredictable amino-acid pairs, and the later become a force driving further mutation, which guarantees continuous protein evolution.

***Corresponding Author:**
Guang Wu
State Key Laboratory of Non-food Biomass Enzyme Technology, National Engineering Research Center for Non-food Biorefinery, Guangxi Key Laboratory of Biorefinery, Guangxi Academy of Sciences
98 Daling Road, Nanning, Guangxi, 530007, China.
E-mail: hongguanglishibahao@yahoo.com

References

[1] Krajcovic M, Overholtzer M. Mechanisms of ploidy increase in human cancers: a new role for cell cannibalism. Cancer Res 2012; 72:1596-601.
[2] Vanstone M, Chow W, Lester L. Ainsworth P, Nisker J, Brackstone M. Recognizing BRCA gene mutation risk subsequent to breast cancer diagnosis in southwestern Ontario. Can Fam Physician 2012; 58:e258-66.
[3] Yamamoto H, Adachi Y, Taniguchi H, Kunimoto H, Nosho K, Suzuki H, Shinomura Y. Interrelationship between microsatellite instability and microRNA in gastrointestinal cancer. World J Gastroenterol 2012;18:2745-55.
[4] Li W, Vijg J. Measuring genome instability in aging - a mini-review. Gerontology 2012; 58: 129-38.
[5] Miller RA. Genes against aging. J Gerontol A Biol Sci Med Sci 2012; 67: 495-502.
[6] Lee HC, Wei YH. Mitochondria and aging. Adv Exp Med Biol 2012;942:311-27.
[7] Cirz RT, Romesberg FE. Controlling mutation: intervening in evolution as a therapeutic strategy. Crit Rev Biochem Mol Biol 2007;42:341-54.
[8] Besaratinia A, Pfeifer GP. Sunlight ultraviolet irradiation and BRAF V600 mutagenesis in human melanoma. Hum Mutat 2008;29:983-91.
[9] Mefford HC, Eichler EE. Duplication hotspots, rare genomic disorders, and common disease. Curr Opin Genet Dev 2009;19:196-204.
[10] Gazzerro E, Sotgia F, Bruno C, Lisanti MP, Minetti C. Caveolinopathies: from the biology of caveolin-3 to human diseases. Eur J Hum Genet 2010;18:137-45.
[11] South ST. Chromosomal structural rearrangements: detection and elucidation of mechanisms using cytogenomic technologies. Clin Lab Med 2011;31:513-24.

[12] Graubert T, Walter MJ. Genetics of myelodysplastic syndromes: new insights. Hematology Am Soc Hematol Educ Program 2011;2011:543-9.
[13] Jena NR. DNA damage by reactive species: Mechanisms, mutation and repair. J Biosci 2012;37:503-17.
[14] Hassen S, Ali N, Chowdhury P. Molecular signaling mechanisms of apoptosis in hereditary non-polyposis colorectal cancer. World J Gastrointest Pathophysiol 2012;3:71-9.
[15] Poincaré H. Science and Hypothesis. Dover Publications: New York, 1952, pp 244.
[16] Wu G, Yan S. Randomness in the primary structure of protein: methods and implications. Mol Biol Today 2002;3:55-69.
[17] Wu G, Yan S. Mutation trend of hemagglutinin of influenza A virus: a review from computational mutation viewpoint. Acta Pharmacol Sin 2006;27:513-26.
[18] Yan S, Wu G. Creation and application of computational mutation. J Guangxi Acad Sci 2010;17:145-50.
[19] Wu G, Yan S. Lecture Notes on Computational Mutation. Nova Science Publishers, New York, 2008.
[20] Yan S, Wu G. Mutation pattern in human adrenoleukodystrophy protein in terms of amino-acid pair predictability. J Biomed Sci Eng 2010;3:262-7.
[21] Yan S, Wu G. Determination of mutation pattern in human androgen receptor by means of amino-acid pair predictability. Protein Pept Lett 2009;16:289-96.
[22] Wu G, Yan S. Determination of amino acid pairs sensitive to variants in human Bruton's tyrosine kinase by means of a random approach. Mol Simul 2003;29:249-54.
[23] Wu G, Yan S. Determination of amino acid pairs sensitive to variants in human coagulation factor IX precursor by means of a random approach. J Biomed Sci 2003;10:451-4.
[24] Wu G, Yan S. Analysis of amino acid pairs sensitive to variants in human collagen α5(IV) chain precursor by means of a random approach. Peptides 2003;24:347-52.
[25] Wu G, Yan S. Amino acid pairs sensitive to variants in human collagen α1 (I) chain precursor. EXCLI J 2004;3:10-9.
[26] Wu G, Yan S. Susceptible amino acid pairs in variants of human collagen α-1(III) chain precursor. EXCLI J 2004;3:20-8.
[27] Wu G, Yan S. Determination of amino acid pairs sensitive to variants in human copper-transporting ATPase 2. Biochem Biophys Res Commun 2004;319:27-31.
[28] Yan S, Wu G. Mutation patterns in human α-galactosidase A. Mol Divers 2010;14:147-54.
[29] Wu G, Yan S. Determination of amino acid pairs sensitive to variants in human β-glucocerebrosidase by means of a random approach. Protein Eng 2003;16:195-9.
[30] Gao N, Yan S, Wu G. Pattern of positions sensitive to mutations in human haemoglobin α-chain. Protein Pept Lett 2006;13:101-7.
[31] Yan SM, Wu G. Mutation pattern of amino-acid pairs in human hemoglobin β-chain. J. Guangxi Acad Sci 2011; 27: 29-33.
[32] Wu G, Yan S. Determination of amino acid pairs sensitive to variants in human low-density lipoprotein receptor precursor by means of a random approach. J Biochem Mol Biol Biophys 2002; 6:401-6.
[33] Yan S, Wu G. Mutation patterns in human menin. Proceedings of 3rd International Conference on Bioinformatics and Biomedical Engineering (iCBBE), 2009, IEEE Xplore ISBN 978-1-4244-2902/8-09.
[34] Yan S, Wu G. Determination of mutation patterns in human ornithine transcarbamylase precursor. J Clin Monit Comput 2009;23:51-7.
[35]. Wu G, Yan S. Determination of amino acid pairs in human p53 protein sensitive to mutations/variants by means of a random approach. J Mol Model 2003;9:337-41.
[36] Wu G, Yan SM. Estimation of amino acid pairs sensitive to variants in human phenylalanine hydroxylase protein by means of a random approach. Peptides 2002;23:2085-90.
[37] Wu G, Yan S. Amino acid pairs susceptible to variants in human protein C precursor. Protein Pept Lett 2005;12:491-4.
[38] Wu G, Yan S. Determination of amino acid pairs in Von Hippel-Lindau disease tumour suppressor (G7 protein) sensitive to variants by means of a random approach. J Appl Res 2003;3:512-20.
[39] Mologni L. Development of RET kinase inhibitors for targeted cancer therapy. Curr Med Chem 2011;18:162-75.
[40] Garcia-Lavandeira M, Diaz-Rodriguez E, Garcia-Rendueles ME, Rodrigues JS, Perez-Romero S, Bravo SB, Alvarez CV. Functional role of the RET dependence receptor, GFRa co-receptors and ligands in the pituitary. Front Horm Res 2010;38:127-38.
[41] Cakir M, Grossman AB. Medullary thyroid cancer: molecular biology and novel molecular therapies. Neuroendocrinology 2009;90:323-48.
[42] Morrison PJ, Atkinson AB. Genetic aspects of familial thyroid cancer. Oncologist 2009;14:571-7.

[43] Salehian B, Liu Z. Genetic alterations in differentiated thyroid cancers. Endocr Metab Immune Disord Drug Targets 2009;9:257-68.
[44] Torino F, Paragliola R.M, Barnabei A, Corsello SM. Medullary thyroid cancer: a promising model for targeted therapy. Curr Mol Med 2010;10:608-25.
[45] Filicori F, Keutgen XM, Buitrago D, AlDailami H, Crowley M, Fahey TJ 3rd, Zarnegar R. Risk stratification of indeterminate thyroid fine-needle aspiration biopsy specimens based on mutation analysis. Surgery 2011;150:1085-91.
[46] Raue F, Frank-Raue K. Genotype-phenotype relationship in multiple endocrine neoplasia type 2. Implications for clinical management. Hormones (Athens) 2009;8:23-8.
[47] Wells SA Jr, Santoro M. Targeting the RET pathway in thyroid cancer. Clin. Cancer Res 2009; 15:7119-23.
[48] Morandi A, Plaza-Menacho I, Isacke CM. RET in breast cancer: functional and therapeutic implications. Trends Mol Med 2011;17:149-57.
[49] Gershon MD. Developmental determinants of the independence and complexity of the enteric nervous system. Trends Neurosci 2010;33:446-56.
[50] Phay JE, Shah MH. Targeting RET receptor tyrosine kinase activation in cancer. Clin Cancer Res 2010;16:5936-41.
[51] Ye L, Santarpia L, Gagel RF. The evolving field of tyrosine kinase inhibitors in the treatment of endocrine tumors. Endocr Rev 2010;4:578-99.
[52] Antonelli A, Fallahi P, Ferrari SM, Mancusi C, Colaci M, Santarpia L, Ferri C. RET TKI: potential role in thyroid cancers. Curr Oncol Rep 2012;14:97-104.
[53]The UniProt Consortium. UniProt Knowledgebase (UniProtKB). 2012; http://www.uniprot.org/uniprot/P07949
[54] Oldenburg J, Schröder J, Brackmann HH, Müller-Reible C, Schwaab R, Tuddenham E. Environmental and genetic factors influencing inhibitor development. Semin Hematol 2004; 41(1 Suppl 1):82-8.
[55] Costers S, Vanhee M, van Breedam W, van Doorsselaere J, Geldhof M, Nauwynck HJ. GP4-specific neutralizing antibodies might be a driving force in PRRSV evolution. Virus Res 2010; 154:104-13.

8/26/2013