

КОЭВОЛЮЦИЯ ДОМЕНОВ КЛЮЧЕВЫХ БЕЛКОВ АПОПТОЗА p53 И MDM2

С.С. Пинтус

Учреждение Российской академии наук Институт цитологии и генетики Сибирского отделения РАН, Новосибирск, Россия, e-mail: pintus@bionet.nsc.ru

Исследованы коэволюционные отношения между доменами белков p53 и mdm2, связанными между собой регуляторными отношениями. Найдены свидетельства коэволюции между отдельными участками взаимодействующих доменов белков. Также оценены скорости аминокислотных замен, предположительно характерные для коэволюционирующих, а также для нейтрально эволюционирующих районов этих белков.

Ключевые слова: p53, mdm2, белок, эволюционное расстояние, коэволюция, скорость замен, пространственная структура.

Введение

Белок p53 относится к супрессорам опухолей. Он способен как останавливать деление клеток, так и вызывать их апоптоз по митохондриальному пути. Белок p53 участвует в передаче клеточного сигнала посредством как активации транскрипции генов-мишеней, так и белок-белковых взаимодействий. Для выполнения функции фактора транскрипции белку p53 необходим трансактивационный домен (TAD), но связывание p53 с регуляторной областью гена-мишени осуществляется посредством домена связывания с ДНК (DBD) (Levine, 1997).

Транскрипционная активность белка p53 в клетке подавляется посредством связывания домена MDM2 убиквитин-лигазы mdm2 с трансактивационным доменом p53. Также это приводит к уменьшению концентрации p53 в клетке за счет убиквитинилирования. В свою очередь, ген *mdm2* является мишенью белка p53, который активирует транскрипцию собственного ингибитора. Таким образом, осуществляется механизм обратной связи, который позволяет поддерживать крайне низкую концентрацию p53 в нормальной клетке (Prives, 1998).

Известно, что эволюционная история белка p53 была связана с его специализацией на функции «стража генома» (Yang *et al.*, 2002), что

подразумевает приобретение им новых генов мишеней и соответственно адаптивные изменения в аминокислотной последовательности трансактивационного домена (TAD). Учитывая важность белок-белкового взаимодействия между TAD-доменом белка p53 и MDM2-доменом белка mdm2, можно предположить, что в истории этих белков имела место коэволюция между трансактивационным доменом p53 и MDM2-доменом mdm2. Таким образом, целью настоящей работы было исследовать коэволюционные отношения между доменами белков p53 и mdm2.

Исследование коэволюции на молекулярном уровне интенсивно ведется в последние годы. Обзор данной области доступен в работе Galtier, Dutheil (2007).

Материалы и методы

Современные методы исследования коэволюции белков основаны на анализе геномных последовательностей и представлены двумя группами. Первая группа подходов основывается на методе анализа корреляции между заменами. Данный метод был впервые предложен в работе Goh *et al.* (2000) и развит в работе Pazos, Valencia (2001). Другая группа методов основана на сходстве топологий филогенетических деревьев (Jothi *et al.*, 2005).

В настоящей работе был использован подход корреляции матриц аминокислотных расстояний для оценки коэволюции белков, аналогичный подходу, предложенному группой Пазоса в их методе поиска взаимодействий между белками на основании их коэволюции (Pazos, Valencia, 2001). Однако в этом подходе не были учтены различия между аминокислотными сайтами с различными скоростями замен.

Подход, основанный на сходстве топологий филогенетических деревьев, представляется менее предпочтительным, поскольку он требует надежной реконструкции филогении. Надежность реконструкции зависит, в частности, от числа сайтов выравнивания, использованного для реконструкции. В настоящей работе были использованы выборки из сравнительно небольшого числа сайтов аминокислотного выравнивания. При исследовании коэволюции между этими выборками снижение надежности филогении оказало бы свое влияние, и на этом основании был выбран подход корреляции матриц аминокислотных расстояний.

Ортологичные аминокислотные последовательности белков семейств p53 и mdm2 были взяты из базы данных Ensembl (Curwen *et al.*, 2004) (www.ensembl.org) для различных организмов, за исключением курицы (*G. gallus*), для которой последовательности были взяты из базы данных Swiss-Prot (Bairoch *et al.*, 2005) (www.uniprot.org/uniprot). Последовательности были отобраны таким образом, чтобы набор представленных видов организмов был одним и тем же для белков p53 и mdm2 (табл. 1).

Из базы данных PFAM (pfam.sanger.ac.uk) были взяты границы доменов TAD, DBD и MDM2 для видов, у которых последовательности белков p53 и mdm2 представлены в базе данных Swiss-Prot. С помощью программы ClustalW 2 были получены выравнивания аминокислотных последовательностей белков p53 и mdm2. При выравнивании были использованы параметры по умолчанию: весовая матрица Gonnet (Gonnet, 2000), штраф на открытие разрыва – 10, штраф на продолжение разрыва – 0,20. Для mdm2 показатель идентичности последо-

Таблица 1

Список видов организмов и идентификаторы белковых последовательностей белков p53 и mdm2 в базе данных Ensembl

Вид организма	Идентификатор аминокислотной последовательности p53	Идентификатор аминокислотной последовательности mdm2
<i>Canis familiaris</i>	ENSCAFP00000024579	ENSCAFP0000000608
<i>Danio rerio</i>	ENSDARP00000051548	ENSDARP00000091133
<i>Equus caballus</i>	ENSECAP00000006210	ENSECAP00000011666
<i>Gallus gallus</i>	P53_CHICK	MDM2_CHICK
<i>Gasterosteus aculeatus</i>	ENSGACP00000025614	ENSGACP00000001415
<i>Homo sapiens</i>	ENSP00000269305	ENSP00000258148
<i>Macaca mulatta</i>	ENSMMUP00000011334	ENSMMUP00000018642
<i>Mus musculus</i>	ENSMUSP00000005371	ENSMUSP00000020408
<i>Ochotona princeps</i>	ENSOPRP00000005906	ENSOPRP00000008864
<i>Oryzias latipes</i>	ENSORLP00000008029	ENSORLP00000021473
<i>Pan troglodytes</i>	ENSPTRP00000014836	ENSPTRP00000050082
<i>Pongo pygmaeus</i>	ENSPPYP00000008923	ENSPPYP00000005415
<i>Rattus norvegicus</i>	ENSRNOP000000047840	ENSRNOP00000008388
<i>Takifugu rubripes</i>	ENSTRUP000000030646	ENSTRUP00000020873
<i>Xenopus tropicalis</i>	ENSXETP000000053759	ENSXETP00000003026

Примечание. Для курицы (*G. gallus*) приведены идентификаторы из базы Swiss-Prot.

вательностей (identity) оказался равным 69/558 (12 %), показатель сходства (similarity) – 100/558 (18 %). Для р53 показатели идентичности и сходства составили 81/424 (19 %) и 116/424 (27 %) (согласно ClustalW 2).

Из полученных выравниваний с учетом известных границ доменов были извлечены столбцы, соответствующие доменам TAD и DBD для р53 и домену MDM2 для mdm2. В случае домена MDM2 области выравниваний границ доменов были отредактированы вручную с учетом аминокислотных последовательностей границ доменов, аннотированных в базах Swiss-Prot и Pfam. В частности, известно, что домен MDM2 довольно консервативен (Bennett-Lovsey *et al.*, 2002), тогда как выравнивание, полученное с помощью ClustalW 2, содержало длинную

последовательность разрывов, соответствующую обширной делеции у человека и макаки, в результате чего консервативная С-концевая последовательность домена у этих организмов смещается (рис. 1, а). Разрывы были удалены вручную таким образом, чтобы С-концевые последовательности совпадали (рис. 1, б).

Из полученных выравниваний были извлечены и выравнены заново фрагменты, соответствующие доменам TAD, DBD и MDM2 (ClustalW 2, параметры по умолчанию). Число сайтов в каждом из выравниваний составило 53, 193, и 81 соответственно (рис. 2).

Три полученных выравнивания были поданы на вход программы PhyML, использовалась дискретная модель гамма-распределения замен, число категорий – 4 (согласно рекомендации ав-

а	
Chicken	VKEHRR IYS <u>MI SRNL</u> IAINQQDSTLAVPPEDDAKFRLEKENVLKESMQELEEKQTSSN--ATSQPTTSRR
Chimpanzee	----KIYT <u>MI YRNL</u> VVVNQEQESSDGSVSENCHLEGGSDQKDLVQELQEEKPSSSHLVS RPSTSSRR
Dog	VKEHRKIYT <u>MI YRNL</u> VVVNQHEPDSGTSVSENCHREGGSDQKDPVQELQEEKPSSDLISRPSTSSRR
Fugu	VKEPRVLFAM <u>ITMKNL</u> VAVKNPESSSQSDSNTDDQR--EKQEEPATPEEEELQEQRSS----LSPHWRSS
Horse	VKEHRKIYT <u>MI YRNL</u> VVVNQPEPDSGTSVSENCPLEGGSDQKAAVQELQEEKPSSSALVSRPPTSSRR
Human	-----VLFYLGQYIMTKRLYDEKQQHIVYCSNDLLG--DLFGVPSFSVKEHRKIYT <u>MI YR</u> ---- <u>NL</u>
Macaque	-----VLFYLGQYIMTKRLYDEKQQHIVYCSNDLLG--DLFGVPSFSVKEHRKIYT <u>MI YR</u> ---- <u>NL</u>
Medaka	VKEPRVLFAM <u>ITKKNL</u> VAVNKAAVNSNKEVKKNKYELEEMLKTL-----LIKKKSS--VKLQMEHQLVA
Mouse	VKEHRKIYA <u>MI YRNL</u> VAVSQQ--DSGTSLSERRQPEGGSDLKDP LQAPPEEKPSDDLISRLSTSSRR
Orangutan	VKEHRKIYT <u>MI YRNL</u> VVVNQEQESSDGSVSENCHLEGGSDQKDLVQELQEEKPSSSHLVS RPSTSSRR
Pika	VKEHRKIYT <u>MI YRNL</u> VVVNQEQEPSDSGMSISEDYHPEGGNDEKEPQELQEEKPSSSLVSRPSTSSRR
Rat	VKEHRKIYA <u>MI YRNL</u> VVVSQQ--DSGTSPESESRCPPEGGSDLKDPVQASQEEKPSSSDVVS RPSTSSRR
Stickleback	VKEPRVLFAM <u>ITKKNL</u> VAVKSQEPASSSEAAAGASRTEEAESSESHSPERRGRKRRRRSRFRSSSEPSGSS
X.tropicalis	VKEPRRLYA <u>MI SRNL</u> VSAAVKESSEVYGVNHCSPFDKQKSKELLQELP-EKVIAPAYDSKPCNSSQR
Zebrafish	VKEPRALFAL <u>INRNL</u> VTVKNPESQSTFSEPRSQSEPD RGP GD TDS-----DSRSS-----TSQQQRRR
б	
Chicken	MKEVIFYLGQYIMSKQLYDEKEQHIVHCANDLLGDLFGVTSFSVKEHRR IYS <u>MI SRNL</u>
Chimpanzee	MKE-----KIYT <u>MI YRNL</u>
Dog	MKEVIFYLGQYIMTKRLYDEKQQHIVYCSNDLLGDLFGVPSFSVKEHRKIYT <u>MI YRNL</u>
Fugu	MKEVMFYLGQYI IQKQLYDQKQQHIVHCSHDELGRVLGVDSFVKEPRVLFAM <u>ITMKNL</u>
Horse	MKEVIFYLGQYIMTKRLYDERQQHVMVYCSNDLLGDLFGVPSFSVKEHRKIYT <u>MI YRNL</u>
Human	MKEVLFYLGQYIMTKRLYDEKQQHIVYCSNDLLGDLFGVPSFSVKEHRKIYT <u>MI YRNL</u>
Macaque	MKEVLFYLGQYIMTKRLYDEKQQHIVYCSNDLLGDLFGVPSFSVKEHRKIYT <u>MI YRNL</u>
Medaka	MKEVMFYLGQYIMQKQLYDQKQQHVMVHCAQDALGRVLGVDSFVKEPRVLFAM <u>ITKKNL</u>
Mouse	MKEIIFYIGQYIMTKRLYDEKQQHIVYCSNDLLGDLFGVPSFSVKEHRKIYA <u>MI YRNL</u>
Orangutan	MKEVLFYLGQYIMTKRLYDEKQQHIVYCSNDLLGDLFGVPSFSVKEHRKIYT <u>MI YRNL</u>
Pika	MKEVIFYLGQYIMTKRLYDEK--HIVYCSND--LGDLFGVPSFSVKEHRKIYT <u>MI YRNL</u>
Rat	MKEIIFYIGQYIMTKRLYDEKQQHIVYCSNDLLGDLFGVPSFSVKEHRKIYA <u>MI YRNL</u>
Stickleback	MKEVMFYLGQYIILKQLYDQKQQHIVHCSQDALGRVLGVDSFVKEPRVLFAM <u>ITKKNL</u>
X.tropicalis	MKEVIFYLGQYIMAKQLYDEKQQHIVHCSNDPLGELFGVQEFVKEPRRLYA <u>MI SRNL</u>
Zebrafish	MKEVMFYLGKYIMSKELYDKQQHIVHCGEDPLGAVLVGKVSFVKEPRALFAL <u>INRNL</u>

Рис. 1. Фрагмент выравнивания последовательностей белка mdm2 до и после ручной правки.

а – выравнивание до ручной правки. Остатки С-концевой последовательности домена MDM2 выделены. Видно смещение положения С-концевой последовательности у человека (Human) и макаки (Macaque). б – фрагмент выравнивания последовательностей mdm2 после ручной правки.

TAD	
14	53
Chicken	-----MAEEMEP-LLEP---TEVFMDLWSMLP-----
Chimpanzee	-----MEEPQSDPSVEPPLSQETFSDLWKLLPENNVLSPL---
Dog	-----MQEPQSELNIDPPLSQETFSELNLLPENNVLSSE---
Fugu	MED---EGFS-----LPLSQDTFQDLWENVML-----
Human	-----MEEPQSDPSVEPPLSQETFSDLWKLLPENNVLSPL---
Macaque	-----MEEPQSDPSIEPPLSQETFSDLWKLLPENNVLSPL---
Medaka	-----WESMDPVP--DLPESQGSFQELWETVY-----
Mouse	-----MEESQSDISLELPLSQETFSGLWKLLPPEDILP-----
Orangutan	-----MEEPQSDPSVEPPLSQETFSDLWKLLPENNVLSPL---
Pika	-----MEEQPSELSIEPPLSQETFPELWNLPL---LLSTN---
Rat	-----MEDSQSDMSIELPLSQETFSCWLKLLPPDDILPTTATG
Sticklebac	MEGTGFEGFLALGQ--DLPDSQDSFAELWGNVVS-----
X.tropical	-----ME-PSSETGMEPPLSQETFEDLWLLP-----
Zebrafish	-----MAQ-----NDSQE-FAELWEKNLI-----

DBD	
15	193
Chicken	DYGGDFDFRV GFVEAGTAKS VTCTYSPVLN KVYCRLAKPC PVQVRVGVAP
Chimpanzee	TYQGSYGFRL GFLHSGTAKS VTCTYSPALN KMFCQLAKTC PVQLWVDSTP
Dog	TYPGTYGFRL GFLHSGTAKS VTWTYSPLLN KLFQQLAKTC PVQLWVSSPP
Fugu	DYPGEYGFEL RFQKSGTAKS VTSTYSEILN KLYCQLAKTS LVEVLLIKKP
Horse	-----YSPTLN KLFQQLAKTC PVQLLVSSPP
Human	TYQGSYGFRL GFLHSGTAKS VTCTYSPALN KMFCQLAKTC PVQLWVDSTP
Macaque	TYHGSYGFRL GFLHSGTAKS VTCTYSPDLN KMFCQLAKTC PVQLWVDSTP
Medaka	DYPGSYELEL RFQKSGTAKS VTSTYSETLN KLYCQLAKTS PIEVRVSKEP
Mouse	TYQGSYGFRL GFLQSGTAKS VMCTYSPPLN KLFQQLAKTC PVQLWVSATP
Orangutan	TYQGSYGFRL GFLHSGTAKS VTCTYSPALN KMFCQLAKTC PVQLWVDSTP
Pika	TYPGTYGFRL GFLQSGTAKS VTCTYSPSLN KLFQQLAKTC PVQLKVDSTP
Rat	TYQGSYGFRL GFLQSGTAKS VMCTYSPSLN KLFQQLAKTC PVQLWVTSTP
Sticklebac	DHPGEFGFQL RFQKSGTAKS VTSTYSEILN KLYCQLAKTS LIEVLVTKEL
X.tropical	DYAGSYGLKL EFQQNGTAKS VTCTYSTDNLN KLFQQLAKTC PLLVVRVERP
Zebrafish	DYPGDHGFRL RFPQSGTAKS VTCTYSPDLN KLFQQLAKTC PVQMVDVAP

MDM2	
15	81
Chimpanzee	LVRPKPLLLK LLKSVGAQKD TYTMKE----
Horse	LVRPKPLLLK LLKSVGAQKD TYTMKEVIFY LGQYIMTKRL YDERQQHMVY
Human	LVRPKPLLLK LLKSVGAQKD TYTMKEVIFY LGQYIMTKRL YDEKQQHIVY
Macaque	LVRPKPLLLK LLKSVGAQKD TYTMKEVIFY LGQYIMTKRL YDEKQQHIVY
Orangutan	LVRPKPLLLK LLKSVGAQKD TYTMKEVIFY LGQYIMTKRL YDEKQQHIVY
Dog	LVRPKPLLLK LLKSVGAQKD TYTMKEVIFY LGQYIMTKRL YDEKQQHIVY
Pika	LVRPKPLLLK LLKSVGAQKD TYTMKEVIFY LGQYIMTKRL YDEK--HIVY
Mouse	LVRPKPLLLK LLKSVGAQND TYTMKEIIFY IGQYIMTKRL YDEKQQHIVY
Rat	LVRPKPLLLK LLKSVGAQKD IYTMKEIIFY IGQYIMTKRL YDEKQQHIVY
Chicken	LVRPKPLLLK LLKLAGAEKD TFTMKEVIFY LGQYIMSKQL YDEKEQHIVH
X_tropical	LVRPKPLLLS LLKSAGAQKE TFTMKEVIYH LGQYIMAKQL YDEKQQHIVH
Medaka	LVRPKVEFCS LLQHAGATKD VFTMKEVMFY LGRYIMQKQL YDQKQQHMVH
Sticklebac	LVQPKLEFRT LLHQAGATKE VFTMKEVMFY LGQYIILKQL YDQKQQHIVH
Fugu	QVRPNVEFHR LLQDAGAIKD VFTMKEVMFY LGQYIIQKQL YDQKQQHIVH
Zebrafish	-VRPKVQLKS LLEDAGADKD VFTMKEVMFY LGKYIMSKEL YDKQQQHIVH

Рис. 2. Фрагменты выравниваний доменов TAD, DBD и MDM2.

торов метода), параметр гамма-распределения (α) и число инвариантных сайтов подбирались программой автоматически, использовалась матрица аминокислотных замен WAG (Whelan, Goldman, 2001), значение бутстреп – 500. В результате были получены три филогенетических дерева (рис. 3).

Построенные деревья, а также полученные выравнивания доменов были использованы для измерения скоростей замен в аминокислотных сайтах выравниваний с помощью программы CODEML пакета PAML 3.14. Аминокислотные сайты каждого домена были разбиты программой на 5 категорий в зависимости от скорости

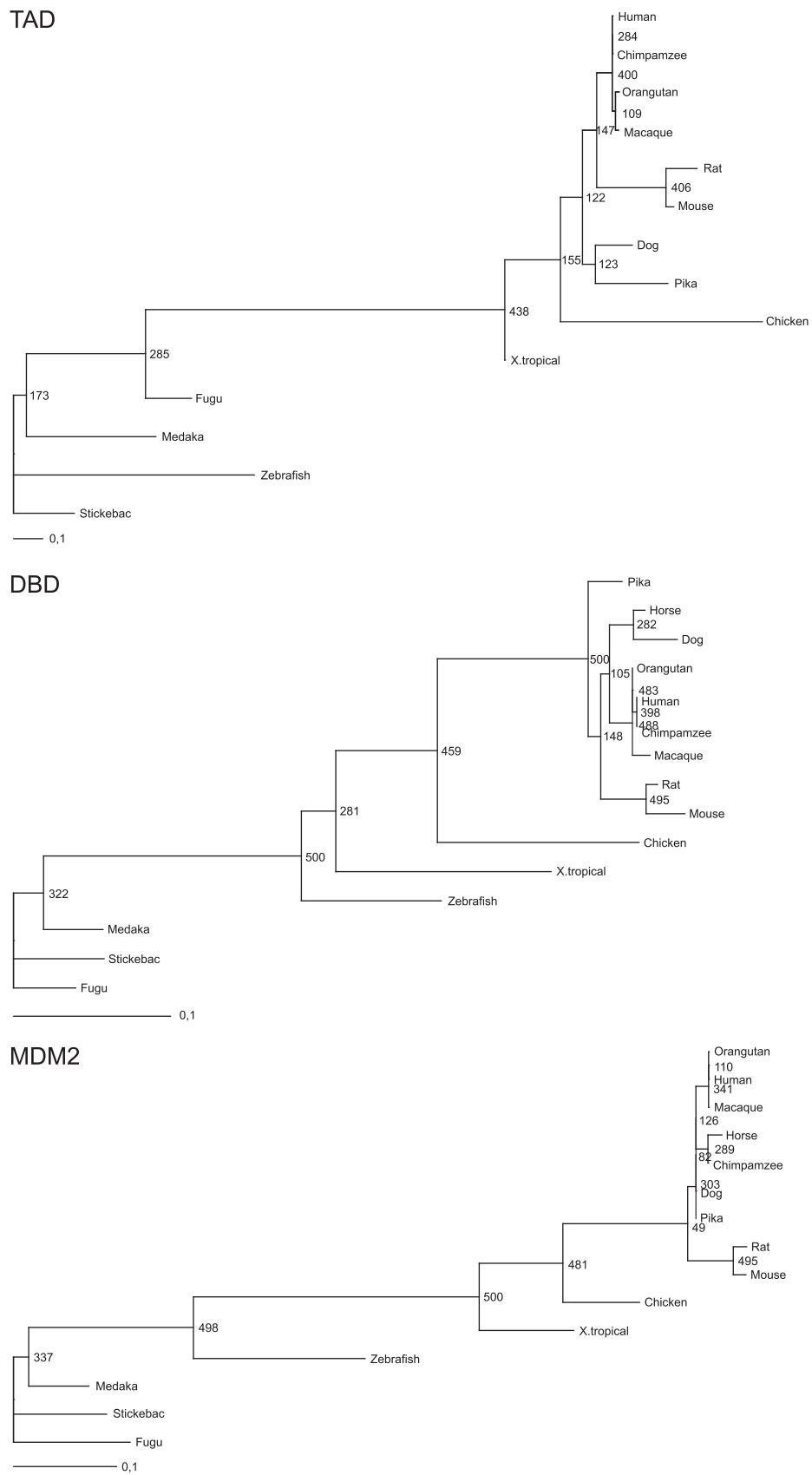


Рис. 3. Филогенетические деревья доменов TAD, DBD и MDM2, построенные программой PhyML.

Числа обозначают значения бутстрепа.

замен. Средние скорости замен в категориях составили 0,02, 0,16, 0,47, 1,11 и 3,25 млн лет⁻¹. При этом использовалась матрица аминокислотных замен WAG (Whelan, Goldman, 2001), параметр гамма-распределения (α) определялся программой, итерации подбора параметра начинались с 0,5.

Выбор количества категорий определялся представленностью различных значений скоростей замен, с одной стороны, и количеством сайтов в категории, достаточным для статистического исследования (построения матрицы корреляции), с другой стороны. Так, при выборе 4 категорий разбиение сайтов программой CODEML таково, что не представлена отдельная группа сайтов со средней скоростью от 1–2 млн лет⁻¹, имеющая довольно большое значение для интерпретации результатов настоящей работы, как поясняется ниже. Однако при выборе 6 категорий в категориях со средними скоростями 0,10 и 0,31 млн лет⁻¹ не представлено ни одного аминокислотного сайта домена MDM2.

Аминокислотные сайты выравнивания каждого из трех доменов были разбиты на 5 групп в зависимости от категории средней скорости замен, к которой они были отнесены программой CODEML. Каждая такая группа составила отдельную выборку (всего 15 выборок). Для каждой выборки была построена матрица эволюционных расстояний с помощью программы LAPD, разработанной Л. Арвстадом (Arvestad, 2006). Между матрицами расстояний, соответствующими одинаковым группам скоростей замен, вычислялся коэффициент попарной корреляции Пирсона, который использовался в качестве меры коэволюции соответствующих участков белков.

Результаты

Частоты встречаемости сайтов (в процентах), относящихся к различным категориям скорости замен, приведены в табл. 2. Можно видеть, что трансактивационный домен р53 отличался от двух других исследованных доменов большей вариабельностью, а также тем, что лишь у него встречались слабовариабельные аминокислотные сайты (0,16 млн лет⁻¹). Интересно, что доля наиболее вариабельных сайтов была примерно одинакова у всех трех доменов, тогда как доли

Таблица 2
Частоты встречаемости аминокислотных сайтов (в процентах от общего числа сайтов домена), относящихся к различным категориям скорости замен

Категория сайтов	Скорость замен, млн лет ⁻¹	mdm2 MDM2	p53 DBD	p53 TAD
1	0,02	0,43	0,47	0,23
2	0,16	0,00	0,00	0,15
3	0,47	0,09	0,11	0,25
4	1,11	0,27	0,22	0,17
5	3,25	0,21	0,20	0,21

полностью консервативных сайтов (0,02 млн лет⁻¹) существенно различались. Также стоит отметить, что распределения сайтов по скоростям замен схожи у доменов mdm2 MDM2 и р53 DBD, тогда как распределение для домена р53 TAD существенно отличается от первых двух большей равномерностью (рис. 4).

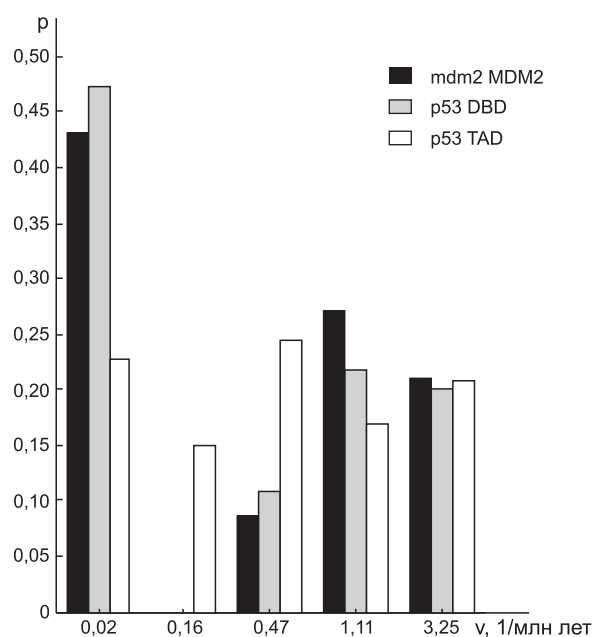


Рис. 4. Распределение частот встречаемости аминокислотных сайтов (в процентах от общего числа сайтов домена), относящихся к различным категориям скорости замен.

По оси абсцисс отложены средние скорости аминокислотных замен для каждой категории (млн лет⁻¹), по оси ординат – частота встречаемости аминокислотных сайтов данной категории.

Значения коэффициента корреляции Пирсона между матрицами эволюционных расстояний в парах доменов белков представлены в табл. 3.

Данные для категории 1 отсутствуют, так как в нее попадали полностью консервативные сайты, и все три матрицы эволюционных расстояний для этой категории состояли из нулей. Ввиду отсутствия сайтов, относящихся к категории 2 у доменов mdm2 MDM2 и p53 DBD, не было возможности оценить коэффициент корреляции и для них.

Как видно из табл. 3, распределение коэффициентов корреляции между матрицами эволюционных расстояний для пары невзаимодействующих доменов mdm2 MDM2 – p53 DBD существенно отличается от таковых для пар взаимодействующих доменов mdm2 MDM2 – p53 TAD и p53 DBD – p53 TAD. Корреляция между высоковариабельными сайтами из категории 5 довольно высока для пары mdm2 MDM2 – p53 DBD и низка для остальных двух пар. Корреляция же между относительно консервативными сайтами из категории 3 наоборот особенно низка для пары mdm2 MDM2 – p53 DBD. Следует также отметить, что для всех трех пар доменов высока корреляция между сайтами из категории 4.

Обсуждение

Отличие распределения сайтов по категориям скоростей замен для транскрипционного домена p53 от распределения для двух других исследованных доменов может отражать события движущего отбора в ходе эволюционной истории белка p53, связанные с приобретением новых генов-мишеней в процессе специализации p53 на функции «стража генома».

Ранее было неоднократно показано, что уровень филогенетического сигнала тем ниже,

чем больше сила естественного отбора (Hansen, Martins, 1996; Diniz-Filho, 2001; Martins *et al.*, 2002; Jose *et al.*, 2008). Для сайтов, несущих самый низкий уровень филогенетического сигнала, можно предположить самые низкие значения корреляции между невзаимодействующими доменами MDM2 и DBD, самые высокие – между постоянно взаимодействующими доменами одного белка p53 и помежуточные – для эпизодически взаимодействующих доменов MDM2 и TAD. Такое распределение значений корреляции наблюдается только для сайтов из категории 3 (0,47 млн лет⁻¹, табл. 3). Отсюда можно сделать вывод, что сайты со скоростью замен порядка 0,47 млн лет⁻¹ несут наименьший уровень филогенетического сигнала в белках p53 и mdm2. Можно предположить, что фиксация замен именно в этих сайтах регулировалась отбором, связанным с коэволюцией этих белков.

В пространственной структуре домена DBD человека сайтам категории 3 соответствуют 20 остатков, преимущественно расположенных на границах между неструктурным и структурированным элементами (8 на границах с β -листами, 3 – на границе с петлями и еще 1 на границе с α -спиралью). Внутри β -листов располагаются 3 наиболее консервативных остатка из 3-й категории. Еще 2 остатка располагаются внутри α -спиралей. Они менее консервативны. Остальные остатки, относящиеся к категории 3, располагаются в неструктурных элементах.

Преобладание в 3-й категории остатков, располагающихся на границах структурных элементов, а значит, важных для пространственной структуры домена, дает дополнительный аргумент в пользу отбора по остаткам из категории 3.

В пространственной структуре α -спирали домена TAD, взаимодействующей с MDM2 (Kussie *et al.*, 1996), остатки категории 3 (S20, D21, K24, L25, L26) прилегают к консерватив-

Таблица 3

Значения коэффициента корреляции Пирсона между матрицами эволюционных расстояний в парах доменов белков mdm2 и p53

Категория сайтов	Скорость замен, млн лет ⁻¹	MDM2 – TAD	DBD – TAD	MDM2 – DBD
3	0,47	0,390 ± 0,004	0,719 ± 0,002	0,207 ± 0,004
4	1,11	0,526 ± 0,004	0,710 ± 0,002	0,730 ± 0,002
5	3,25	0,187 ± 0,005	0,220 ± 0,005	0,825 ± 0,001

ным остаткам F19, T23, P27, взаимодействующим с MDM2. Остаток P27 также взаимодействует с MDM2, но относится к 3-й категории.

В пространственной структуре домена MDM2 остатки, соответствующие сайтам категории 3 (V52, Q71, I74, S90, M102), располагаются вблизи остатков L54, Q72, H73, V75, F91, T100, взаимодействующих с TAD. Более того, остатки I74, S90, а также K31 располагаются на границах β -листов аналогично домену DBD.

Такое распределение сайтов категории 3 на пространственной структуре взаимодействующих доменов TAD и MDM2 позволяет сделать вывод, что остатки, непосредственно участвующие во взаимодействии белков p53 и mdm2, консервативны, тогда как по аминокислотным сайтам белков, оказывающим непосредственное влияние на положение взаимодействующих остатков и относящимся к 3-й категории по значению скорости замен, возможен заметный естественный отбор.

Одинаково высокие значения корреляции между сайтами из категории 4 свидетельствуют о том, что сайты из данной категории несут наибольший уровень филогенетического сигнала. Это позволяет предположить, что в данную категорию попали сайты, эволюционирующие нейтрально. Это подтверждается довольно высокой средней скоростью замен в сайтах данной категории – 1,11 млн лет⁻¹. С другой стороны, для сайтов из категории 4 значение корреляции между доменами MDM2 и TAD не так высоко, как в парах доменов MDM2-DBD и DBD-TAD.

Это можно объяснить относительно невысоким уровнем филогенетического сигнала, который несут аминокислотные сайты домена TAD. В пользу этого предположения говорит то, что именно в домене TAD категория 3 наиболее представлена (рис. 4, табл. 2). Другим аргументом является разброс длин ветвей филогенетического древа домена TAD, гораздо больший, чем для доменов MDM2 и DBD. Разброс длин ветвей свидетельствует о том, что различия скоростей эволюции в разных таксонах у домена TAD были наибольшими из всех трех доменов. Следовательно, для домена TAD характерен наименьший филогенетический сигнал.

Особый интерес представляют низкие значения корреляции сайтов категории 5 для пар взаимодействующих доменов mdm2 MDM2 –

p53 TAD и p53 DBD – p53 TAD и высокое значение для невзаимодействующих доменов MDM2. Из 39 аминокислотных сайтов DBD человека, относящихся к 5-й категории, 35 располагаются в неструктурных элементах, соединяющих структурные элементы. Два других – L206 и V217 – располагаются в центре β -листов «бочонка», являющегося структурным ядром домена. Если сайт V217 инвариантен у всех таксонов, за исключением трех видов рыб, у которых в этом сайте наблюдаются остатки треонина, цистеина и фенилаланина, то сайт L206 весьма вариабелен. Учитывая важность структуры «бочонка», образованного β -листами, для функции всего домена, можно предположить, что наблюдаемые высокие скорости замен в некоторых сайтах 5-й категории связаны не с нейтральной эволюцией, а с отбором. Скорость замен в части сайтов данной категории, следовательно, можно считать переоцененной, что может быть связано с недостатком последовательностей p53 промежуточных таксонов и вытекающей из этого переоценкой вариабельности остатков в сайтах этой категории. Этот вывод подтверждается тем, что 15 из 17 сайтов 5-й категории на структуре домена MDM2 человека располагаются в структурных элементах, однако преимущественно удалены от интерфейса взаимодействия доменов.

Из распределения сайтов 5-й категории на пространственной структуре доменов MDM2 и P53 можно предположить, что высокое значение корреляции для пары невзаимодействующих доменов MDM2 и DBD обусловлено большой долей нейтрально эволюционирующих сайтов в белке p53, попавших в 5-ю категорию, но несущих филогенетический сигнал. Благодаря невысокой доле неструктурных элементов в доменах MDM2 и TAD, сайты 5-й категории из этих доменов оказались свободны от филогенетического сигнала, благодаря чему не наблюдается высокой корреляции для этой пары доменов. Низкие значения корреляции для этих сайтов в парах доменов MDM2-TAD и DBD-TAD можно объяснить отбором, связанным не с взаимодействием доменов, а с поддержанием структурной целостности каждого домена. Это косвенно подтверждается удаленностью сайтов 5-й категории в MDM2 от интерфейса взаимодействия.

Работа поддержана междисциплинарным интеграционным проектом фундаментальных исследований СО РАН № 26 «Математические модели, численные методы и параллельные алгоритмы для решения больших задач СО РАН и их реализация на многопроцессорных суперЭВМ», а также междисциплинарным интеграционным проектом фундаментальных исследований СО РАН № 113 «Разработка вычислительных методов, алгоритмов и аппаратурно-программного инструментария параллельного моделирования природных процессов».

Литература

- Arvestad L. Efficient methods for estimating amino acid replacement rates // *J. Mol. Evol.* 2006. V. 62. P. 663–673.
- Bairoch A., Apweiler R., Wu C.H. *et al.* The Universal Protein Resource (UniProt) // *Nucl. Acids Res.* 2005. V. 33. P. D154–D159.
- Bennett-Lovsey R., Hart S.E., Shirai H., Mizuguchi K. The SWIB and the MDM2 domains are homologous and share a common fold // *Bioinformatics.* 2002. V. 18. P. 626–630.
- Curwen V., Eyraas E., Andrews T.D. *et al.* The Ensembl automatic gene annotation system // *Genome Res.* 2004. V. 14. P. 942–950.
- Diniz-Filho J.A. Phylogenetic autocorrelation under distinct evolutionary processes // *Evolution.* 2001. V. 55. P. 1104–1109.
- Galtier N., Duthel J. Coevolution within and between genes // *Genome Dyn.* 2007. V. 3. P. 1–12.
- Goh C.S., Bogan A.A., Joachimiak M. *et al.* Co-evolution of proteins with their interaction partners // *J. Mol. Biol.* 2000. V. 299. P. 283–293.
- Gonnet G.H., Korostensky C., Benner S. Evaluation measures of multiple sequence alignments // *J. Comput. Biol.* 2000. V. 7. P. 261–276.
- Hansen T.F., Martins E. Translating between microevolutionary process and macroevolutionary patterns: the correlation structure of interspecific data // *Evolution.* 1996. V. 50. P. 1404–1417.
- José J., Puma-Villanueva W.J., Von Zuben F.J., Diniz-Filho J.A. Phylogenetic inheritance of genetic variability produced by neutral models of evolution // *Genet. Mol. Res.* 2008. V. 7. P. 1327–1343.
- Jothi R., Kann M.G., Przytycka T.M. Predicting protein-protein interaction by searching evolutionary tree automorphism space // *Bioinformatics.* 2005. V. 21. P. i241–i250.
- Kussie P.H., Gorina S., Marechal V. *et al.* Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain // *Science.* 1996. V. 274. P. 948–953.
- Levine A.J. p53, the cellular gatekeeper for growth and division // *Cell.* 1997. V. 88. P. 323–331.
- Martins E.P., Diniz-Filho J.A., Housworth E.A. Adaptive constraints and the phylogenetic comparative method: a computer simulation test // *Evolution.* 2002. V. 56. P. 1–13.
- Pazos F., Valencia A. Similarity of phylogenetic trees as indicator of protein-protein interaction // *Protein Eng.* 2001. V. 14. P. 609–614.
- Prives C. Signaling to p53: breaking the MDM2-p53 circuit // *Cell.* 1998. V. 95. P. 5–8.
- Whelan S., Goldman N. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach // *Mol. Biol. Evol.* 2001. V. 18. P. 691–699.
- Yang A., Kaghad M., Caput D., McKeon F. On the shoulders of giants: p63, p73 and the rise of p53 // *Trends Genet.* 2002. V. 18. P. 90–95.

COEVOLUTION OF PROTEIN DOMAINS OF P53 AND MDM2 KEY PROTEINS OF APOPTOSIS

S.S. Pintus

Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia, e-mail: pintus@bionet.nsc.ru

Summary

Coevolution of domains of p53 and mdm2 proteins, that are involved in mutual regulation, is studied. Evidences for coevolution of particular regions of interacting domains of these two proteins are found. Amino acid substitution rates are estimated for regions of the proteins that are supposed to undergo coevolution as well as for regions of the proteins that are supposed to undergo neutral selection.