

ГЕНОМИКА.
ТРАНСКРИПТОМИКА

УДК 577.21

**ВНУТРИМОЛЕКУЛЯРНАЯ ГОМОЛОГИЧНАЯ РЕКОМБИНАЦИЯ В ГЕНЕ
БЕЛКА ШЕЛКА ЯЙЦЕВОГО КОКОНА ПАУКА-ОСЫ *Argiope bruennichi*[#]**

© 2013 г. Leng Han, Masao Nakagaki*

Division of Applied Biology, Faculty of Textile Science and Technology, Shinshu University, Ueda, Nagano 386-8567, Japan

Поступила в редакцию 18.01.2013 г.

Принята к печати 18.04.2013 г.

С целью изучения семейства генов белков шелка яйцевых коконов в 2006 г. Zhao и соавт. выделили две полных последовательности кДНК, кодирующих цилиндрические белки шелка 1 (CySp1) и 2 (CySp2) паука-осы *Argiope bruennichi*. Согласно данным этой работы, в CySp2 отсутствует характерная последовательность сигнального пептида, а комплекс CySp1—CySp2, который содержит сигнальный пептид, должен поступать в эндоплазматический ретикулум и секретироваться в комплекс Гольджи. Секвенирование геномной ДНК — один из подходов, который можно успешно применять для получения 5'-концевых участков генов белков шелка. Используя этот метод, мы выделили 5'-концевую область гена CySp2. Оказалось, что CySp2 содержит типичный сигнальный пептид, сходный с обнаруженным в CySp1. Таким образом, последовательность CySp2, определенная Zhao и соавт., содержит ошибку, обусловленную техническими сложностями секвенирования генов белков шелка.

Ключевые слова: CySp2, мутация сдвига рамки, сигнальный пептид, внутримолекулярная гомологичная рекомбинация.

INTRAMOLECULAR HOMOLOGOUS RECOMBINATION EVENT OCCURRED IN THE SPIDER EGG CASE SILK GENE CySp2 OF WASP SPIDER *Argiope bruennichi*[#] by Leng Han, Masao Nakagaki* (Division of Applied Biology, Faculty of Textile Science and Technology, Shinshu University, Ueda, Nagano 386-8567, Japan; *e-mail: nakgak@shinshu-u.ac.jp). To gain further understanding of egg case silk proteins gene family, Zhao et al. isolated two full-length cDNAs for egg case silk proteins, cylindrical silk protein 1 (CySp1) and cylindrical silk protein 2 (CySp2), from the wasp spider, *Argiope bruennichi*. CySp2 was reported to contain no apparent signal peptide sequences, and the CySp1-CySp2 complex, which would possess a signal peptide, would be transported across the endoplasmic reticulum and secreted to the Golgi. Genomic DNA sequencing is one approach that can be successfully utilized to retrieve 5' ends of silk genes; using this method, we retrieved the 5' end of CySp2. We found that CySp2 contained a typical signal peptide similar to that found in CySp1; thus, due to technical limitations, an artificial error had occurred in the CySp2 sequence reported by Zhao et al.

Keywords: CySp2, frameshift mutation, signal peptide, intramolecular homologous recombination.

DOI: 10.7868/S002689841305011X

Шелк паука (белковый полимер) привлекает интерес ученых, благодаря своим уникальным механическим свойствам. Шелк многих пауков состоит из необычных белков, кодируемых большими генами, длиной примерно 10000 п.н. Поскольку гены белков шелка имеют очень большой размер, получение полных нуклеотидных последовательностей их кДНК представляет сложную задачу.

Развитие методов секвенирования ускорило идентификацию новых генов белков шелка пауков. Некоторые из определенных последовательностей соответствуют мРНК, тогда как другие соответствуют геномной ДНК. Однако большин-

ство из них представляют собой лишь частичные кодирующие последовательности. К настоящему времени получены полные кодирующие последовательности таких генов белков шелка паука, как гены главных пузырьковых спидроинов 1 и 2 *Latrodectus hesperus* (*LhMaSp1*, EF595246 и *LhMaSp2*, EF595245 соответственно) и гены цилиндрического белка 2 шелка *Argiope bruennichi* (*AbCySp1*, AB242144 и *AbCySp2*, AB 242145 соответственно) [1, 2]. В настоящей работе определена геномная последовательность 5'-концевой области гена CySp2. В отличие от Zhao и соавт. [2] мы обнаружили, что CySp2 кодирует характерный сигнальный пептид, следова-

[#] Текст представлен авторами на английском языке.

*Эл. почта: nakgak@shinshu-u.ac.jp

тельно, гипотеза, согласно которой транспорт комплекса CySp1—CySp2 через эндоплазматический ретикулум зависит только от сигнального пептида CySp1, неверна.

ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

Содержание пауков и выделение геномной ДНК.

Зрелых самок пауков *A. bruennichi* отлавливали в г. Уено (префектура Нагано, Япония) в октябре. Геномную ДНК выделяли из 25 мг мозга с использованием набора DNeasy Blood & Tissue kit (“QIAGEN”).

Полимеразная цепная реакция (ПЦР) и выделение из геля. ПЦР проводили на приборе ABI Thermal Cycler 2700 по следующей программе: начальная денатурация — 94°C, 5 мин, затем 28 циклов (денатурация — 94°C, 30 с; отжиг — 57°C, 30 с; удлинение праймера — 72°C, 30 с) и конечная дотройка при 72°C, 10 мин. ПЦР проводили со следующей парой праймеров: прямой праймер — 5'-GGATGCCATTGCCCAGGTGATAC-3' и обратный праймер — 5'-AAACGCCCTGACCGGCTAAGAAT-3'. После электрофореза полосы ДНК вырезали из агарозного геля чистым бритвенным лезвием и очищали, используя набор Wizard® SV Gel и PCR Clean-Up System (“Promega”).

Секвенирование и анализ ДНК. Очищенные фрагменты клонировали в Т-векторе pMD19 (“TaKaRa”, Япония), используя набор DNA Ligation Kit Ver.2.1 (“TaKaRa”). Вставки в полученном векторе секвенировали с двух сторон с универсальных праймеров — M13 primer RV и M13 primer M4 (“TaKaRa”) — на секвенаторе ABI PRISM® 3100 Genetic Analyzer. Компьютерный анализ нуклеотидных и аминокислотных последовательностей проводили, используя программу Genetux ver.10. Новую N-концевую последовательность CySp2 (CySp2N) анализировали с помощью программы signalP4.0 (eukaryotic option) [3]. Сравнение опубликованной [2] N-концевой последовательности CySp2 (CySp2O) и CySp2N проводили, используя программы Blast P и Blast N, на сервере NCBI. Выравнивание транскрибируемой нуклеотидной последовательности CySp2O и определенной нами геномной последовательности CySp2N выполняли online, используя SIM4 [4].

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

Получение геномной последовательности, кодирующей N-концевую область CySp2

Ранее секвенирование геномной ДНК использовали для поиска 5'-концов генов белков шелка. Поскольку некоторые гены шелка паука лишены интронов [1], мы выбрали метод секвенирования геномной ДНК для определения 5'-концевых последовательностей гена *CySp2*. 5'-Концевая область гена *CySp2* обладает большим сходством с

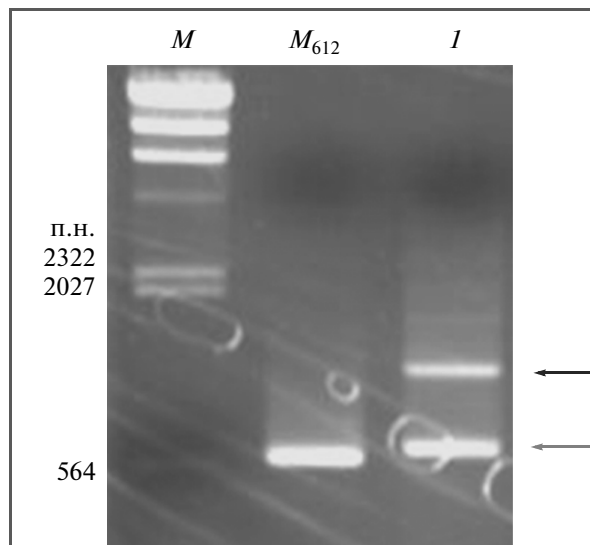


Рис. 1. ПЦР-амплификация геномной ДНК для получения 5'-концевого фрагмента *CySp2*. *M* — HindIII-перевар ДНК фага λ , стандартный маркер молекулярных масс (“TaKaRa”). *M*₆₁₂ — фрагмент ДНК 612 п.н. *I* — Продукт ПЦР-амплификации геномной ДНК. Нижняя стрелка указывает на фрагмент гена *CySp2*, от 5'-концевой неповторяющейся области до первой повторяющейся единицы. Верхняя стрелка указывает на фрагмент гена, который включает 5'-концевую неповторяющуюся область и несколько повторяющихся единиц (возможно, две).

CySp1 (идентичность — 271/292, 93%, результаты выравнивания не показаны). Примечательно, что *CySp2* и *CySp1* построены преимущественно из гомогенных сборных повторов [2], поэтому мы сконструировали набор праймеров, специфичных для *CySp2*. Прямой праймер — 5'-GGATGCCATTGCCCAGGTGATAC-3', локализован в неповторяющейся 5'-области гена, а обратный — 5'-AAACGCCCTGACCGGCTAAGAAT-3', находится в первой единице повтора *CySp2O*. ПЦР-продукт геномной ДНК размером 650 п.н. был амплифицирован (отмечен стрелкой на рис. 1).

Секвенирование и компьютерный анализ фрагмента ДНК

Небольшой фрагмент ДНК (на рис. 1 показан стрелкой) был очищен из агарозного геля и клонирован в Т-векторе для секвенирования. Обнаружено, что фрагмент состоит из 655 п.н.¹ (см. дополнительный материал 1a, Sl1a www.molecbio.com/downloads/2013/5/supp_length_en.pdf). Объединив фрагмент 655 п.н. с последовательностью, представленной Zhao и соавт. [2], получили новую геномную последовательность 5'-области гена *CySp2* — *CySp2N*

¹ Ссылки на дополнительные материалы, представленные на сайте журнала, отмечены в тексте буквой S.

Гомология N-концевых последовательностей *CySp1* и *CySp2N*

Сравнение с *CySp2O* показало, что *CySp2N* содержит дополнительные 349 п.н. Выравнивание N-концевых последовательностей *CySp2N* и *CySp1* показало, что две неповторяющиеся N-концевые области высокогомологичны на уровне как нуклеотидной (идентичность – 497/534, 93%, S4a), так и аминокислотной последовательности (идентичность – 164/178, 92%, S4b).

ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Согласно ранее опубликованным данным, гены белков шелка паука либо состоят из одного экзона [1], либо имеют экзон-интронную структуру [5, 6]. *CySp2N* отличается от *CySp2O* наличием дополнительного фрагмента из 349 п.н. *CySp2N* – это геномная последовательность. Для проверки возможности того, что делетированный фрагмент принадлежит интронной области, с использованием программы SIM4 проанализированы две последовательности (*CySp2O* и *CySp2N*). Показано, что предполагаемый сайт сплайсинга содержит пары СТ–ГС (выделены серым на рис. 2). Однако структура всех известных сайтов сплайсинга генов белков шелка удовлетворяет золотому правилу GT–AG [5]. Учитывая ограниченное число известных геномных последовательностей генов белков шелка пауков, мы проверили сайты сплайсинга у других видов. Помимо GT–AG (97%) в сайтах сплайсинга выявлено несколько типов неканонических пар нуклеотидов: AT–AG, GA–AG, GT–AC, GT–CG и GT–TG (3%). Идентифицированная в потенциальном сайте сплайсинга *CySp2N* пара “СТ–ГС” не входит в этот набор. По сравнению с *CySp2O*, *CySp2N* содержит дополнительные 349 п.н., высокогомологичных последовательности *CySp1* и других генов белков шелка пауков (данные не приведены). Таким образом, фрагмент 349 п.н. не принадлежит интронной последовательности, он соответствует мРНК, транскрибируемой с геномной ДНК. Иными словами, *CySp2O* не является вариантом сплайсинга [7], возникновение делеции не связано с секвенированием после конструирования библиотеки кДНК (личное сообщение Zhao).

Внутримолекулярные рекомбинации с высокой вероятностью происходят по простым прямым повторам, обычно в ходе конструирования плазмид (неопубликованные данные). Исходя из результатов проведенного анализа (S2b и рис. 3) мы предположили, что акт эксцизии произошел в ходе конструирования библиотеки кДНК. Полученные при обратной транскрипции–ПЦР двухцепочечные фрагменты ДНК встраивали в плазмиду и трансформировали в клетки *Escherichia coli*, при этом в процессе саморепликации плазмиды

произошла внутримолекулярная гомологичная рекомбинация (рис. 3), которая привела к делеции 349 п.н. Мы полагаем, что участок *CySp2O*, состоящий из 7 п.н. (нуклеотиды 321–328), – это след такого события рекомбинации.

Сигналы секреции – ключевые элементы белков шелка пауков и шелкопрядов, важные для секреции этих белков посредством транспорта через эндоплазматический ретикулум [8]. Согласно Zhao и соавт. [2] в *CySp2O* отсутствует сигнальный пептид. В действительности, определенная ими последовательность *CySp2* содержит ошибку, появление которой обусловлено техническими ограничениями и характерными особенностями структуры генов белков шелка пауков (состоящих из сотен прямых повторов). Путем выравнивания нескольких протяженных EST из библиотеки кДНК нами восстановлена полная нуклеотидная последовательность кДНК *CySp2* (9.8 т.п.н.). Кроме того, делеция, идентифицированная в 5'-концевой области гена, расположена между неповторяющейся N-концевой и повторяющейся областью. Более того, размер делетированного фрагмента равен 349 п.н., что некратно 3. Несомненно, что делеция этого фрагмента привела к искусственной мутации сдвига рамки. Если при анализе гена использовать интактную рамку считывания (RF2) после стартового кодона, выбранного Zhao и соавт. [2], то в нижележащей области встретится множество стоп-кодонов. По этой причине *CySp2O* не содержит сигнальной последовательности. Сходные результаты других работ показывают, что гены белков шелка пауков лишены последовательностей, кодирующих сигнальный пептид; это может быть артефактом, обусловленным отсутствием вышележащих участков в некоторых клонных этих генов. Можно полагать, что если 5'-концевая последовательность простирается до области, кодирующей неповторяющийся N-концевой участок, то метод обращенной ПЦР позволит получить полную N-концевую последовательность.

Таким образом, гипотеза, согласно которой N-концевые остатки *Cys* вовлечены в образование связей между *CySp1* и *CySp2* подобно C-концевым остаткам *Cys* в других белках шелка, неверна. Представление о том, что комплекс *CySp1*–*CySp2*, который должен содержать сигнальный пептид, будет затем транспортироваться через эндоплазматический ретикулум и секретироваться в комплекс Гольджи [2], также может быть неверным. Согласно недавним сообщениям N- и C-концевые области играют ключевую роль в запасании и сборке белков шелка и не обладают функцией сигнального пептида [9]. Линии шелкопряда *Nd-s* и *Nd-s^D* продуцируют исключительно мало фиброина не потому, что мутантные гены легких (L) цепей лишены последовательностей, кодирующих сигнальный пептид, а потому, что у мутантов *Nd-s* и *Nd-s^D* не об-

разуется дисульфидная связь между тяжелыми (H) и легкими (L) цепями из-за отсутствия Cys172 в участке мутантной L-цепи, кодируемом 3'-концевой областью гена [10].

Искусственная мутация сдвига рамки в участке, кодирующем N-концевую область CySp2O, полностью меняет ее аминокислотную последовательность. Поэтому ранее [2] ошибочно отмечали, что CySp1 и CySp2 содержат существенно различающиеся N-концевые участки.

Почему ошибки этого типа сохранялись длительное время? По нашему мнению, возможны две причины этого. Первая – это географическая изоляция. Некоторые виды пауков обитают только в нескольких странах. Ученые из других стран не имеют доступа к материалам этих опытов, хотя и могут предполагать ошибочность результатов. Для предотвращения заноса чуждых или экзотических видов во многих странах запрещено ввозить некоторых животных, в том числе пауков. Во-вторых, полные нуклеотидные последовательности генов белков шелка всегда получают путем секвенирования библиотек кДНК или геномных ДНК. Поэтому получение полных последовательностей генов белков шелка паука стоит очень дорого.

В заключение следует сказать, что нами проведен повторный анализ N-концевой области CySp2. Обнаружено, что в результате ошибки, допущенной в работе [2], в последовательности CySp2 делегирован фрагмент из 349 п.н. Показано, что N-концевая область CySp2 содержит сигнальный пептид и высокоомологична N-концевой области CySp1. Более того, между участками, кодирующими неповторяющуюся N-концевую область, и первыми повторами A- и B-типа CySp2 отсутствует интрон.

Работа получили поддержку Министерства образования, культуры, спорта, науки и технологий Японии (Global COE program by the Ministry of Education, Culture, Sports, Science and Technology of Japan).

СПИСОК ЛИТЕРАТУРЫ

1. Ayoub N.A., Garb J.E., Tinghitella R.M., Collin M.A., Hayashi C.Y. 2007. Blueprint for a high-performance biomaterial: full-length spider dragline silk genes. *PLoS One*. **2**, e514.
2. Zhao A.C., Zhao T.F., Nakagaki K., Zhang Y.S., Sima Y.H., Miao Y.G., Shiomi K., Kajiura Z., Nagata Y., Takadera M., Nakagaki M. 2006. Novel molecular and mechanical properties of egg case silk from wasp spider, *Argiope bruennichi*. *Biochemistry*. **45**, 3348–3356.
3. Petersen T.N., Brunak S., von Heijne G., Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Meth.* **8**, 785–786.
4. Florea L., Hartzell G., Zhang Z., Rubin G.M., Miller W. 1998. A computer program for aligning a cDNA sequence with a genomic DNA sequence. *Genome Res.* **8**, 967–974.
5. Motriuk-Smith D., Smith A., Hayashi C.Y., Lewis R.V. 2005. Analysis of the conserved N-terminal domains in major ampullate spider silk proteins. *Biomacromolecules*. **6**, 3152–3159.
6. Hayashi C.Y., Lewis R.V. 2000. Molecular architecture and evolution of a modular spider silk protein gene. *Science*. **287**, 1477–1479.
7. Rising A., Hjalms G., Engstrom W., Johansson J. 2006. N-terminal nonrepetitive domain common to dragline, flagelliform, and cylindrical spider silk proteins. *Biomacromolecules*. **7**, 3120–3124.
8. Hayashi C.Y., Lewis R.V. 1998. Evidence from flagelliform silk cDNA for the structural basis of elasticity and modular nature of spider silks. *J. Mol. Biol.* **275**, 773–784.
9. Eisoldt L., Thamm C., Scheibel T. 2012. Review the role of terminal domains during storage and assembly of spider silk proteins. *Biopolymers*. **97**, 355–361.
10. Inoue S., Tanaka K., Tanaka H., Ohtomo K., Kanda T., Imamura M., Quan G.X., Kojima K., Yamashita T., Nakajima T., Taira H., Tamura T., Mizuno S. 2004. Assembly of the silk fibroin elementary unit in endoplasmic reticulum and a role of L-chain for protection of alpha1,2-mannose residues in N-linked oligosaccharide chains of fibrohexamerin/P25. *Eur. J. Biochem./FEBS*. **271**, 356–366.