

Evolución de la familia de las globinas en Deuterostomados: Patrones linaje-específicos de diversificación y desgaste

Federico G. Hoffmann,^{*,1,2} Juan C. Opazo,³ David Hoogewijs,⁴ Thomas Hankeln,⁵ Bettina Ebner,⁵ Serge N. Vinogradov,⁶ Xavier Bailly,⁷ and Jay F. Storz⁸ ¹Department of Biochemistry and Molecular Biology, Mississippi State University ²Institute for Genomics, Biocomputing and Biotechnology, Mississippi State University ³Instituto de Ecología y Evolución, Facultad de Ciencias, Universidad Austral de Chile, Valdivia, Chile ⁴Institute of Physiology and Zürich Center for Integrative Human Physiology, University of Zürich, Zürich, Switzerland ⁵Institute of Molecular Genetics, Johannes Gutenberg-University, Mainz, Germany ⁶Department of Biochemistry and Molecular Biology, Wayne State University School of Medicine ⁷Station Biologique de Roscoff, Roscoff, France ⁸School of Biological Sciences, University of Nebraska ^{*}Corresponding author: E-mail: federico.g.hoffmann@gmail.com. Associate editor: Yoko Satta

Abstract

En los Metazoarios, las proteínas de la familia de las globinas muestran una unidad subyacente en la estructura terciaria que contrasta con una extraordinaria diversidad de estructuras primarias, propiedades bioquímicas y funciones fisiológicas. Las reconstrucciones filogenéticas pueden revelar cuáles de estas funciones representan innovaciones novedosas y específicas del linaje, y cuáles representan funciones ancestrales que se comparten con las globinas homólogas en otros eucariotas e incluso procariotas. Hasta la fecha, nuestra comprensión de la diversidad de la globina en los deuteróstomos se ha visto obstaculizada por la escasez de datos de secuencias genómicas de Ambulacraria (equinodermos y hemichordados), el grupo hermano de los cordados, y el filo Xenacoelomorpha, que incluye a los xenoturbélidos, los acelomorfos y las nemertodermátides. Aquí se presentan los resultados de un análisis filogenético y genómico comparativo del repertorio de genes de globinas de los deuteróstomos. Primero caracterizamos los genes de globinas del gusano bellota, *Saccoglossus kowalevskii*, un representante del filo Hemichordata. Luego integramos los datos de la secuencia genómica del gusano bellota en un análisis exhaustivo de las relaciones filogenéticas y sinténicas conservadas entre los genes de las globinas de los representantes de los ocho linajes que componen el superfamilia Deuterostomia. Los objetivos principales fueron 1) desentrañar la historia evolutiva de la superfamilia de los genes de las globinas en los deuterostomados y 2) utilizar la filogenia estimada para obtener información sobre la evolución funcional de las globinas de deuterostomados. Los resultados de nuestros análisis indican que el ancestro común de deuterostomados poseía un repertorio de al menos cuatro parálogos de globinas distintos y que diferentes subconjuntos de estos genes ancestrales han sido retenidos en cada uno de los linajes de los organismos descendientes. En cada uno de los principales grupos de deuteróstomos, un subconjunto diferente de genes precursores ancestrales sufrió expansiones de la diversidad funcional específicas del linaje, a través de repetidas rondas de duplicación y divergencia de genes. Al integrar los resultados del análisis filogenético con los datos funcionales disponibles, descubrimos que las hemoglobinas de transporte de oxígeno circulantes evolucionaron de forma independiente en varios linajes deuteróstomos y que las globinas nerviosas intracelulares evolucionaron de forma independiente en los cordados y los gusanos acoelomorfos.

Introducción

La unidad estructural de las globinas animales se refleja en el altamente conservado "pliegue de globinas", un característico pliegue tridimensional de seis a ocho hélices que encierra un grupo hemo no unido covalentemente en una bolsa hidrofóbica. Esta característica estructural diagnóstica está altamente conservada entre globinas homólogas de todos los reinos de la vida a pesar de la extensa divergencia en la secuencia de aminoácidos (Lesk y Chothia 1980; Lecomte et al. 2005). En los animales, las globinas se expresan en una amplia gama de tejidos y tipos de células, incluidas las células musculares y nerviosas, así como los glóbulos rojos circulantes, y en muchos invertebrados, las globinas extracelulares se encuentran en fluidos vasculares, celómicos y/o perientéricos (Weber y Vinogradov 2001). Esta diversidad en los sitios anatómicos y celulares de expresión está asociada con una variedad correspondientemente amplia de estructuras cuaternarias. Las globinas animales funcionan como monómeros de un solo dominio, proteínas de un solo dominio de varias subunidades, o proteínas de varias subunidades de varios dominios con 2-18 dominios de globinas vinculadas covalentemente por cadena (Vinogradov 1985; Riggs 1991; Vinogradov et al. 1993; Terwilliger 1998; Weber y Vinogradov 2001; Proyecto-García et al. 2010). Algunas de las elaboraciones más notables de la estructura cuaternaria se encuentran en las globinas extracelulares de los invertebrados. La mioglobina monomérica (Mb; aprox 17 kDa) y la hemoglobina tetramérica (Hb; aprox 64 kDa) de los vertebrados gnatostomados son eclipsados por las globinas de múltiples subunidades "hexagonales de bicapa" de los gusanos anélidos y vestimentíferos (aprox 3,600 kDa), que se componen de 144 cadenas de globinas

y 36 péptidos linker (Lamy et al. 1996), y las globinas multidominio y de múltiples subunidades, de moluscos bivalvos (;800-12.000 kDa) (Terwilliger y Terwilliger 1978). La diversidad de las estructuras primarias y cuaternarias mostrada por globinas animales subyace en una diversidad igualmente rica de función, que va desde la unión reversible de oxígeno que es central a la familiar función de transporte y almacenamiento de oxígeno de las Hb y Mb de vertebrados, así como de las funciones relacionadas con facilitar la difusión de oxígeno, su detección, la búsqueda de especies reactivas de oxígeno y nitrógeno, señalización redox, y funciones enzimáticas que involucran actividades NO oxigenasa y reductasa (Weber y Vinogradov 2001; Pesce et al. 2002; Fago et al. 2004; Hankeln et al. 2005; Burmester y Hankeln 2009). El análisis filogenético es necesario para determinar cuáles de estas funciones representan innovaciones linaje-específicas, y que representan funciones ancestrales que se comparten con las proteínas globinas homólogas en otros eucariotas e incluso procariotas (Hardison 1996, 1998; Vinogradov et al. 2005, 2006, 2007; Vinogradov y Moens 2008; Kakar et al. 2010). Las funciones principales de muchas globinas animales aún no han sido dilucidadas, y una mirada a sus orígenes evolutivos y afinidades filogenéticas pueden sugerir hipótesis acerca de sus roles fisiológicos. Aunque la mayoría de las globinas de vertebrados son productos de eventos de duplicación de genes o de todo el genoma que ocurrieron en el linaje madre de los vertebrados (Hoffmann et al. 2011, 2012; Storz et al. 2011), dos globinas monoméricas, la neuroglobina (Ngb) y la globina X (GbX), tienen una historia más antigua, ya que se derivan de las duplicaciones que son anteriores al origen de

los deuterostomados (Roesner et al.2005; Ebner et al. 2010). La gran antigüedad y la amplia distribución geográfica de estas dos globinas sugiere que pueden estar realizando

las globinas tenían una distribución filética limitada entre los deuterostomados (Vinogradov et al. 2006; Vinogradov y Moens 2008). Un número de globinas de transporte de oxígeno

Table 1. Globin Genes Identified in the Acorn Worm Genome Assembly.

Label	Protein Record	Gene ID	Number of Codons	Exons	Genomic Location
Acorn worm <i>Gb1</i>	NP_001161601	LOC100313670	155	3	scaffold_29426
Acorn worm <i>Gb2</i>	XP_002732485	LOC100378901	183	3	scaffold_9907
Acorn worm <i>Gb3</i>	XP_002732488	LOC100366547	151	3	scaffold_9907
Acorn worm <i>Gb4</i>	XP_002732489	LOC100366699	151	3	scaffold_9907
Acorn worm <i>Gb5</i>	XP_002732490	LOC100366851	224	4	scaffold_9907
Acorn worm <i>Gb6</i>	XP_002733371	LOC100374916	183	3	scaffold_14411
Acorn worm <i>Gb7</i>	XP_002739222	LOC100375093	204	3	scaffold_38908
Acorn worm <i>Gb8</i>	XP_002739227	LOC100375836	243	3	scaffold_38908
Acorn worm <i>Gb9</i>	XP_002739228	LOC100375983	180	3	scaffold_38908
Acorn worm <i>Gb10</i>	XP_002739229	LOC100376130	163	3	scaffold_38908
Acorn worm <i>Gb11</i>	N/A	N/A	193	1	scaffold_25907
Acorn worm <i>Gb12</i>	N/A	N/A	189	1	scaffold_38407
Acorn worm <i>Gb13</i>	N/A	N/A	191	1	scaffold_44414
Acorn worm <i>Gb14</i>	N/A	N/A	200	1	scaffold_3524
Acorn worm <i>Gb15</i>	N/A	N/A	242	3	scaffold_38908
Acorn worm <i>Gb16</i>	N/A	N/A	174	3	scaffold_38908

NOTE.—Acorn worm *Gbs* 11–16 were manually annotated and do not have an associated protein record or gene ID number. N/A, not applicable.

funciones fisiológicas relacionadas con aspectos muy fundamentales del metabolismo celular. El caso de *GbX* es particularmente interesante porque ha sido recientemente reportado que *GbX* es una proteína asociada a la membrana que puede tener una función antioxidante (Blank et al. 2011). Así, Identificación de ortólogos *Ngb* y *GbX* en taxones de deuteróstomos no cordados pueden proporcionar información adicional sobre las funciones fisiológicas de estas proteínas. A pesar de los avances significativos en nuestra comprensión de evolución de la familia de genes de la globinas en los cordados (Ebner et al. 2003, 2010; Hoffmann, Opazo, et al. 2010, 2011, 2012; Hoffmann, Storz, et al. 2010; Storz et al. 2011), nuestro entendimiento de la diversidad de las globinas en los deuteróstomos en su conjunto ha sido proporcionado por datos de secuencias genómicas para *Ambulacraria* (equinodermos y hemicordados), el grupo hermano de los cordados, y el filo *Xenacoelomorpha*, que incluye a los xenoturbélidos, acelomorfos y nemertodermatites. Originalmente se creía que

han sido caracterizadas funcionalmente en equinodermos, incluyendo el pepinos de mar (clase *Holothuroidea*) y la clase *Ophiuroidea* (Terwilliger y Read 1972; Bonaventura et al. 1976; Suzuki 1989; Mauri et al. 1991, McDonald et al. 1992; Baker and Terwilliger 1993; Kitto et al. 1998, Christensen et al. 2003), y búsquedas bioinformáticas en el genoma del erizo de mar (clase *Echinoidea*) revelaron la presencia de un gen de globina de 34 exones que codifica un solo polipéptido con 16 dominios de globina unidos covalentemente (Bailly y Vinogradov 2008). Aquí, reportamos los resultados de un análisis filogenético y genómico comparativo de el conjunto de genes de globinas de los deuteróstomos. Como primer paso, se caracterizó el repertorio genético de las globinas del gusano bellota, *Saccoglossus kowalevskii*, un representante del phylum *Hemichordata*. A continuación, se integraron datos genéticos de secuencia del gusano en un análisis comprensivo de sintenia conservada y relaciones filogenéticas, entre los genes de representantes de los ocho principales linajes que componen el superfilo *Deuterostomia*. Los

objetivos principales fueron: 1) desentrañar la historia evolutiva de la superfamilia génica de las globinas en deuterostomados y 2) para utilizar la filogenia estimada para comprender la evolución funcional de las globinas de deuterostomados. Los resultados de nuestros análisis indican que el ancestro común de los deuterostomados poseía un repertorio de al menos cuatro parálogos de globinas distintos y que diferentes subconjuntos de estos genes ancestrales han sido retenidos en cada uno de los linajes de organismos descendientes que sobreviven hasta el día de hoy. La distribución filética de los genes de globinas entre los taxones de deuterostomados es claramente atribuible a una avalancha de diversidad ancestral tal que diferentes sets de parálogos estructuralmente distintos han sido retenidos en diferentes linajes, y cada uno de estos parálogos sembró diferentes expansiones linaje-específicas de diversidad funcional a través de repetidas rondas de duplicación de genes y divergencia.

Materiales y Métodos

Análisis bioinformáticos

Utilizamos herramientas bioinformáticas para ensamblar un conjunto de datos que combina los repertorios completos de genes de globina de 18 taxones de deuterostomas, incluidos 16 cordados (13 vertebrados, 2 urocordados y 1 cefalocordado) más 1 hemicordado y 1 equinodermo como representantes de Ambulacraria. Las especies de vertebrados incluyen cinco peces teleósteos (fugu, *Takifugu rubripes*; medaka, *Oryzias latipes*; pez ferf, *Tetraodon nigroviridis*; espinoso de tres espinas, *Gasterosteus aculeatus*; y pez cebra, *Danio rerio*), un anfibio (rana con garras occidental, *Xenopus tropicalis*), un reptil cuadrado (lagartija verde de anole, *Anolis carolinensis*), tres pájaros (pollo, *Gallus gallus*; pavo, *Meleagris gallopavo*; y pinzón cebra, *Taeniopygia guttata*), y tres mamíferos (humanos, *Homo sapiens*; zarigüeya

gris de cola corta, *Monodelphis domestica*, y ornitorrinco, *Ornithorhynchus anatinus*). Las especies no vertebradas incluyeron dos chorros de mar (*Ciona intestinalis* y *C. savignyi*, Urochordata), amphioxus (*Branchiostoma floridae*, Cephalochordata), el erizo de mar morado (*Strongylocentrotus purpuratus*, Echinodermata) y el gusano bellota (*S. kowalevskii*), Hemich. Los datos del gusano de la bellota se obtuvieron del Proyecto del genoma del gusano de bellota Hoffmann et al. · Doi: 10.1093/molbev/mss018 MBE1736 organizado por el Centro de Secuencia del Genoma Humano del Baylor College of Medicine. Debido a que el erizo de mar posee un gen de globina que codifica 16 supuestos dominios de globina (Bailey y Vinogradov 2008), consideramos cada uno de los dominios como secuencias de globina alineables por separado en todos los análisis.

Además de recuperar los repertorios completos de genes de globina de las especies enumeradas anteriormente, también incluimos secuencias de globina individuales de varios taxones de deuterostomas adicionales. Para los equinodermos, incluimos secuencias de dos subunidades de los glóbulos rojos celulómericos *Hbs* del pepino de mar, *Caudina arenicola*. Para los peces cartilagosos, obtuvimos secuencias α y β -Hb de la pastinaca roja (*Dasyatis akajei*) y el gomero (*Mustelus antarcticus*), así como las secuencias Mb del gomero y el tiburón Port Jackson (*Heterodontus portusjacksoni*). Para las aves, incluimos la secuencia de globina E de la liberación preliminar del genoma del pato real (*Anas platyrhynchos*), disponible en la base de datos Ensembl-Pre. La mayoría de los vertebrados gnatóstomos incluidos en este estudio poseen múltiples copias paralógicas de genes de globina tipo α y β . Dado que la monofilia de las familias de genes de globina α y β ha sido bien establecida (Goodman et al. 1975, 1987), solo incluimos un subconjunto representativo de las globinas α y β de cada especie en nuestros análisis. En los análisis

genómicos de los diversos taxones cordados y del erizo de mar, nuestras búsquedas bioinformáticas identificaron el mismo complemento de genes de globina que se había anotado previamente. Debido a que no había información disponible sobre el repertorio de globina del genoma del gusano de bellota, utilizamos un conjunto representativo de las globinas deuterostomas previamente caracterizadas para realizar búsquedas bioinformáticas de secuencias homólogas en el genoma de este organismo utilizando los programas BLASTP, PSI-BLAST y TBLASTN

(Altschul et al. 1990, 1997).

Las estimaciones actuales de las relaciones sistemáticas de deuterostomados colocan xenoturbélidos, gusanos acelomorfos y nemertodermátidas en el filo Xenacoelomorpha, que se considera hermana de Ambulacraria (Bourlat et al.2006; Philippe et al.2011). Por lo tanto, para incluir representantes de cada uno de los principales linajes de deuterostomas, obtuvimos dos secuencias de globina del acoel Symsagittifera roscoffensis, tres secuencias de globina del nemertodermátide Nemertoderma

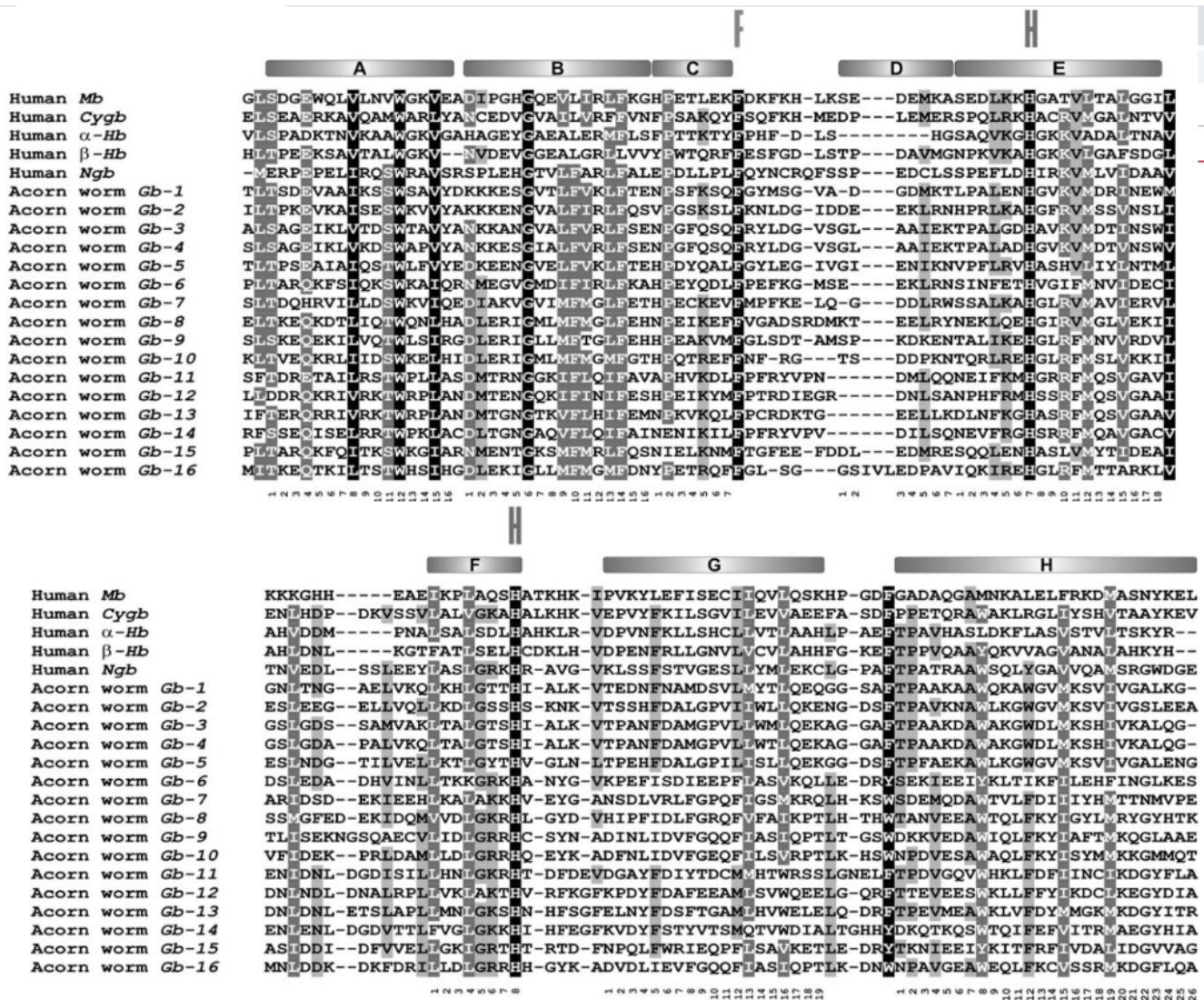


Fig. 1. Structural alignment of acorn worm globins to Human Mb, Cygb, α -Hb, β -Hb, and Ngb sequences. The α -helical structure is shown on top of the alignments. The functionally important phenylalanine in position CD1 and the distal and proximal histidines in positions E7 and F8 are indicated. Numbers correspond to the Human Mb sequence, which was used as a reference. N- and C-terminal extensions of the alignment are not shown.

westbladi, más una secuencia del xenoturbellid *Xenoturbella bocki*. Las globinas de acoel derivan del proyecto del genoma de acoel, mientras que las globinas de nemertodermatida y xenoturbellida se identificaron bioinformáticamente en el Archivo de seguimiento del Centro Nacional de Información Biotecnológica.

La lista completa de secuencias y los números de acceso correspondientes se presentan en la tabla complementaria S1 (Material complementario en línea). Finalmente, con base en los resultados de Hoogewijs et al. (2012), agregamos nueve secuencias de globina de plantas, que se incluyeron como secuencias de grupos externos en análisis filogenéticos. La lista de secuencias de plantas y los números de acceso correspondientes se presentan en la tabla complementaria S2 (Material complementario en línea).

Análisis filogenéticos

Realizamos nuestros análisis filogenéticos en dos etapas. En la primera etapa, alineamos todas las secuencias usando MAFFT con selección de estrategia automática (Katoh y Toh 2008), y usamos reconstrucciones filogenéticas de unión de vecinos como se implementaron en MEGA versión 5 (Tamura et al. 2011) para identificar y eliminar secuencias redundantes. En la segunda etapa, utilizamos el conjunto de datos podados para construir múltiples alineaciones usando una variedad de enfoques: Kalign2 (Lassmann et al. 2009), las estrategias E-INS-i, G-INS-i y L-INS-i de MAFFT v6.83 (Katoh y Toh 2008; Katoh et al. 2009), Muscle (Edgar 2004), PROMALS3d (Pei et al. 2008) y T-coffee (Notredame et al. 2000). Además, también probamos un enfoque de alineación de perfil en el que las secuencias de globina de gusano de bellota, erizo de mar, pepino de mar y *C. savignyi*, además de las de las globinas de la planta se agregaron a una alineación previamente ensamblada de secuencias de globina de vertebrados, anfibios y *C. intestinalis*

(Hoffmann et al. 2011).

Luego empleamos MUMSA (Lassmann y Sonnhammer 2005, 2006) para clasificar las alineaciones múltiples y luego utilizamos el paquete suiteMSA versión 1.2.06 (Anderson et al. 2011) para comparar visualmente entre las dos alineaciones con las puntuaciones MUMSA más altas y seleccionamos la alineación desde el enfoque de alineación de perfil por razones estructurales. Posteriormente, utilizamos la alineación seleccionada para estimar las relaciones filogenéticas utilizando métodos bayesianos implementados en MrBayes versión 3.1.2 (Ronquist y Huelsenbeck 2003) bajo un modelo mixto de sustitución de aminoácidos. Establecimos dos corridas independientes de seis cadenas simultáneas para 10,000,000 generaciones, muestreando cada 2,500 generaciones, y usando los previos predeterminados. Una vez que se verificó la convergencia, el apoyo para los nodos y las estimaciones de los parámetros se derivaron de un consenso de reglas mayoritarias de los últimos 2.500 árboles. Todos los árboles estaban enraizados con plantas Hbs.

Resultados y discusión

El repertorio de genes de globinas de hemicordados

Nuestras encuestas bioinformáticas de bases de datos genómicas y transcriptómicas para el gusano de bellota revelaron la presencia de 17 genes de globina supuestos que tenían un tamaño de 151 a 243 codones (tabla 1). Dieciséis de estas secuencias se derivaron del genoma borrador del gusano de bellota y una se derivó de un registro de ARNm. Debido a que la última secuencia (ACY92586) era idéntica a uno de los registros derivados del genoma (NP_001161601), excluimos la secuencia derivada de ARNm de todos los análisis adicionales. Las 16 secuencias de globina de gusano de bellota se encontraron en ocho andamios genómicos diferentes, y las

predicciones bioinformáticas indicaron que 11 de ellas tienen la estructura 3-exón-2-intrón típica de las globinas específicas de vertebrados, 1 tiene una estructura 4-exón-3-intrón, y los 4 restantes son genes sin intrones.

Las 16 globinas de gusano de bellota identificadas parecen ser globinas de buena fe según varios criterios. Las longitudes de la secuencia de codificación estuvieron dentro del rango esperado para las proteínas de globina funcionales y fueron similares en tamaño a las globinas de los chorros de mar (Ebner et al. 2003) y anfioxo (Ebner et al. 2010). Además, la alineación de la secuencia de aminoácidos reveló que se conservan residuos funcionalmente importantes en todas las secuencias de globina de gusano de bellota (fig. 1). Los sitios conservados incluyen la fenilalanina CD1 y las histidinas proximales (F8) y distales (E7), que están involucradas en la coordinación del grupo hemo y / o la estabilización del ligando Fe²⁺, respectivamente.

Un árbol filogenético de globinas de deuterostomados

Las estimaciones actuales de la filogenia de deuterostomados (véase el árbol inserto en la figura 2) ubican al grupo Ambulacraria, que incluye equinodermos y hemicordados, como hermana del filo Xenacoelomorpha, que incluye cánulas, nemertodermátidas y xenoturbélidos (Bourlat et al. 2006; Philippe et al. 2011). Para reconstruir la historia evolutiva de la superfamilia del gen de la globina deuterostomados, combinamos el conjunto completo de globinas del gusano de la bellota con una amplia gama de globinas de equinodermos en el subfilo Echinozoa (el erizo de mar morado [clase Echinozoidea] y el pepino de mar [clase Holothuroidea]) más representantes de cada uno de los principales cordados y linajes xenacoelomorfos.

El conjunto de datos inicial de 180 secuencias

incluía una sobrerrepresentación de globinas de vertebrados en relación con las de los deuterostomados no vertebrados, así como secuencias redundantes (la lista completa de secuencias y los números de acceso correspondientes se presentan en las tablas complementarias S1 y S2, Material complementario en línea). Para lograr una representación taxonómica más equilibrada sin sacrificar la cobertura filogenética de las diversidades de globinas, primero generamos una alineación de secuencia múltiple usando todas las secuencias y luego construimos un árbol de unión de vecinos (figura complementaria S1, Material complementario en línea).

El árbol resultante nos permitió identificar secuencias redundantes y muy similares, que luego eliminamos para podar nuestro conjunto de datos inicial a 110 secuencias. Luego alineamos estas secuencias con ocho estrategias alternativas (ver Material y Métodos), clasificamos las alineaciones resultantes en base a su puntaje MUMSA y comparamos las dos alineaciones con los puntajes MUMSA más altos (tabla suplementaria S3, Material suplementario en línea) usando el paquete suiteMSA ver 1.2.06 (Anderson et al. 2011). Por razones estructurales, seleccionamos el del enfoque de alineación de perfil para todos los análisis filogenéticos posteriores (archivo de datos complementarios S1, material complementario en línea).

Las filogenias bayesianas organizaron todas las globinas de deuteróstomos distintas de amphioxus Gb8 y las bellotas sin intrón del gusano de bellota (gusano de bellota Gb11, Gb12, Gb13 y Gb14) en cuatro clados confiables (fig. 2), cada uno de ellos definido por la presencia de globinas de cordados y ambulacrarian. El primero de estos grupos, el clado 1, contiene todas las globinas específicas de vertebrados (es decir, Cygb [incluida la Hbs de ciclostomados], Mb, a- y b-Hb, GbE y GbY),

más dos grupos separados de globinas anfioxus, todos urocordados globinas, un clado de globinas de bellota, que es hermana de un clado que incluye nemertodermatid y globinas xenoturbellid, y las dos isoformas de la subunidad Hb del pepino de mar, que son hermanas del acoel Gb1. El segundo grupo, el clado 2, contiene el gusano de bellota Gbs 7–10, 16, acoel Gb2 y el vertebrado GbX y es hermana del tercer grupo, el clado 3, que contiene el gusano de bellota Gb6 y Gb15, y amphioxus Gbs 3, 6, 12–14. Finalmente, el cuarto grupo, el clado 4, incluye los subdominios de erizo de mar, además de vertebrados Ngb y amphioxus Gb4. Las relaciones entre estos cuatro clados están bien resueltas, con el clado 4 que contiene Ngb hermana de los tres clados restantes.

Estas filogenias colocaron a las globinas del gusano de la bellota en cuatro clados separados y colocaron las seis globinas xenacoelomorfos en tres clados separados (fig. 2). Los gusanos de

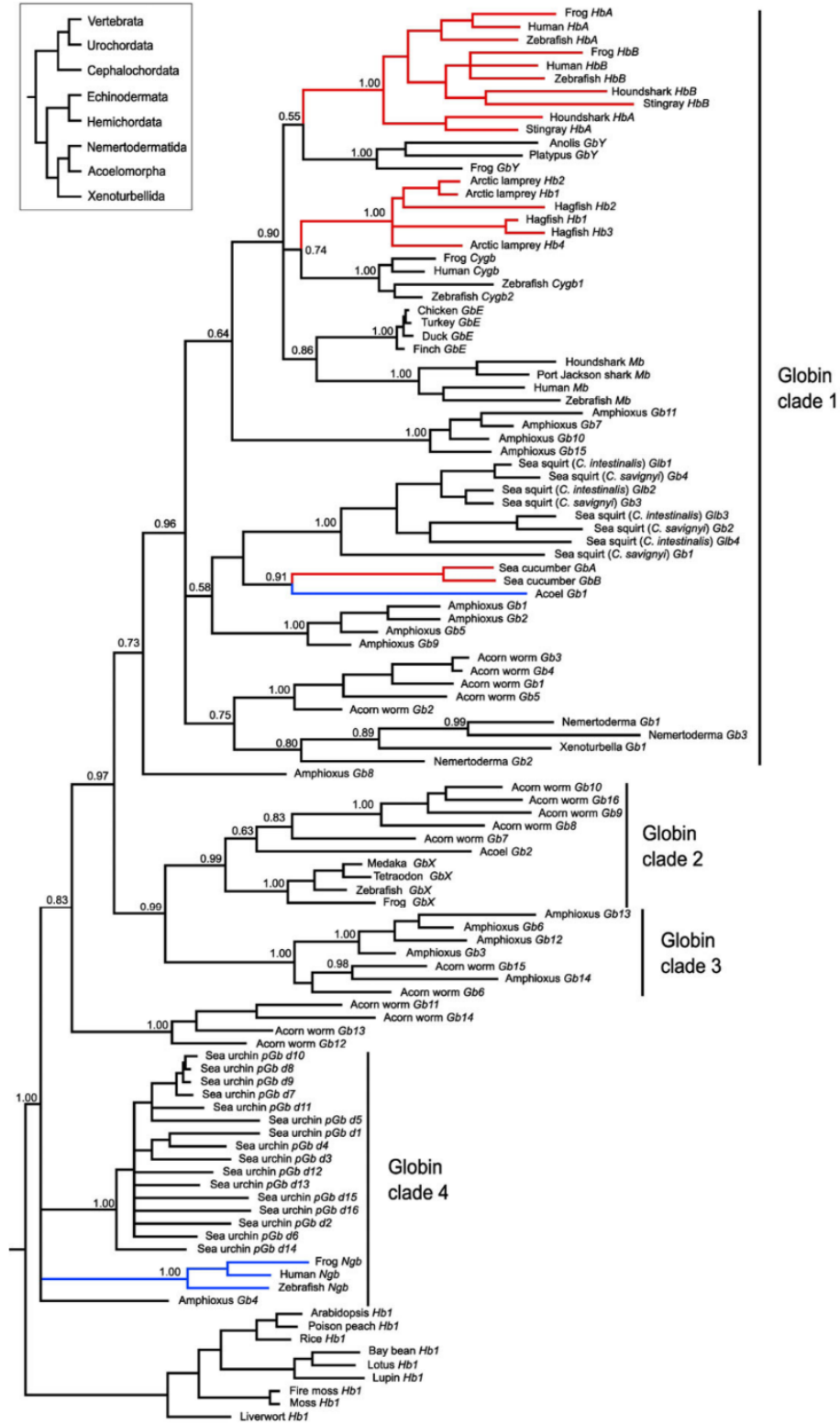
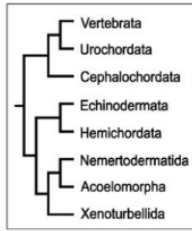
bellota Gb6 y Gb15 se colocaron en clado de globina 3, junto con amphioxus Gbs 3, 6 y 12–14 siendo confiable la relación, y el clado que contenía gusanos de bellota Gbs 7–10 y Gb16 se colocaron en el clado 2, junto con acoel Gb2 y secuencias de vertebrados GbX también siendo fiables. El gusano de bellota Gbs 1–5 se colocó en el clado de globina 1, en un clado hermano de las globinas de Nemertoderma y Xenoturbella, mientras que las globinas de gusano de bellota sin intrón (gusano de bellota Gbs11–14) se colocaron en un clado profundamente divergente con filogenética incierta afinidades Finalmente, el resto de gloel de acoel, acoel Gb1, que es funcionalmente similar al vertebrado Ngb (Bailey X, datos no publicados) también se coloca en el clado de globina 1, pero hermana de las dos isoformas de la subunidad Hb del pepino de mar. En todos los casos, las globinas de xenacoelomorfo se colocan hermanas a las globinas del ambulatorio, como se esperaba

bajo la disposición sistemática que se muestra en el recuadro de la figura 2.

La disposición filogenética en la figura 2 es consistente con estudios previos en la organización de globinas de vertebrados en aquellos que derivan de duplicaciones específicas de vertebrados (Cygb, GbE, GbY, Mb y Hbs de vertebrados con mandíbula y sin mandíbula) y aquellos que derivan de duplicaciones anteriores La divergencia entre deuterostomas y protostomas (GbX y Ngb; Roesner et al. 2005; Hoffmann, Opazo, et al. 2010; Droège y Makalowski 2011; Hoffmann et al. 2011, 2012; Hoogewijs et al. 2012; Storz et al. 2011). Además, como en Hoogewijs et al. (2012), nuestros resultados también identifican al clado que contiene vertebrados Ngb como la división más profunda entre las globinas de deuterostomados.

Patrones de sintenia conservada corroboran las reconstrucciones filogenéticas

Debido a que las globinas son proteínas pequeñas, proveen un número limitado de sitios informativos par resolver relaciones filogenéticas. Por lo tanto, se utilizaron comparaciones de la sintenia conservada para proporcionar una segunda línea de evidencia con respecto a los orígenes duplicativos de las globinas de deuteróstomos. Este análisis de la sintenia conservada reveló que las localizaciones genómicas de las globinas del gusano bellota están en concordancia con las filogenias inferidas. Las cinco globinas del gusano bellota que formaron un clado hermano a GbX (Gbs 7-10, 16) de vertebrados, están dispuestos en tándem a lo largo del scaffold_38908, que también incluye la Gb15 del gusano bellota. Cuatro de los cinco genes del gusano en el clado 1 de las globinas (Gbs 2-5) también están dispuestos en tándem en scaffold_9907, y las globinas de gusano de bellota restantes se encuentran en andamios



separados (tabla 1).

Las comparaciones genómicas de la sintenia conservada brindan soporte adicional para los clados de globinas 2–4 en la filogenia de las globinas de deuterostomados (fig. 3). Por ejemplo, en el caso del clado de la globina 4, los vertebrados *Ngb*, *amphioxus Gb4* y el gen multidominio de la globina del erizo de mar púrpura (*pGb*) se encuentran adyacentes a un ortólogo del gen humano *POMT2* (fig. 3A), que es consistente con los resultados reportados por Ebner et al. (2010). En el caso del clado de globina 2, las cinco globinas del gusano de la bellota que son coortólogos de vertebrados *GbX* están dispuestas en un grupo que está flanqueado por ortólogos del gen *SRP14* humano, que también se encuentra en la región flanqueante de *GbX* en rana, pez cebrá, espinoso, medaka y fugu (fig. 3B). Las comparaciones de sintenia para el clado de globina 3 se complican por el hecho de que los cuatro genes de globina *amphioxus* en este clado se encuentran en scaffolds separados (Ebner et al. 2010), y lo mismo se aplica a las dos globinas de bellota en este clado. Curiosamente, tanto el gen gusano de bellota *Gb6* y el *amphioxus Gb3* se encuentran junto a un ortólogo del gen *SMOC2* humano (fig. 3C). Las comparaciones anteriores de sintenias habían identificado similitudes en el contexto genómico entre vertebrados *GbX* y *amphioxus Gb3*, lo que sugiere que podrían ser ortólogos (Ebner et al. 2010).

Modificaciones postraduccionales en los clados de globinas 2 y 3

Varias globinas del gusano de la bellota contienen extensiones N y C-terminales, como se observa para las globinas de nematodos, moscas de la fruta y *amphioxus*, así como los vertebrados *Cygb* y *GbX* (Burmester y Hankeln 1999; Burmester et al. 2002, 2006; Roesner et al. 2005; Hoogewijs et al. 2008; Ebner et

al. 2010). Curiosamente, la evidencia reciente sugiere que el vertebrado *GbX* es una proteína unida a la membrana, y que los sitios de miristoilación y palmitoilación en la extensión N-terminal de esta proteína son críticos para su correcta localización subcelular (Blank et al. 2011). Tanto la N-miristoilación como la S-palmitoilación son modificaciones postraduccionales características de las proteínas asociadas a la membrana (Resh 1999; Farazi et al. 2001; Linder y Deschenes 2007), y por lo tanto, la presencia de sitios potenciales de N-miristoilación y S-palmitoilación podría proporcionar pistas sobre la localización subcelular y el papel funcional de las proteínas analizadas. Usando herramientas de predicción computacional (*Myristoylator* [Bologna et al. 2004], *MYR Predictor* [Maurer-Stroh et al. 2002] y *CSS-Palm* [Ren et al. 2008]), identificamos un supuesto sitio de miristoilación N-terminal (en Gly2) en diez de las globinas de gusano de bellota y un sitio de palmitoilación (en Cys3) en nueve de ellos (tabla complementaria S4, material complementario en línea). Además, se predijo que todas las secuencias de *GbX* de vertebrados tenían sitios de miristoilación y palmitoilación, como fue el caso de la mayoría de las globinas incluidas en los clados 2 y 3. Estos resultados sugieren que estos genes podrían codificar globinas asociadas a la membrana con funciones fisiológicas similares a la del vertebrado *GbX*.

El repertorio de globinas del antepasado común de deuterostomados

Con el propósito de reconstruir el complemento completo de los genes de la globina en el ancestro común del deuterostomados, razonamos que los grupos monofiléticos que contienen secuencias de representantes de *Ambulacraria* (equinodermos y hemicordados) y *Chordata* (vertebrados, cefalocordados y urocordados) deberían rastrear su ascendencia

hasta el nodo ancestral de los deuterostomados. De acuerdo con este criterio, inferimos que el último ancestro común de los deuterostomados existentes poseía un repertorio de al menos cuatro genes de globina parálogos. Estos cuatro genes de globina

de linajes de globina adicionales que podrían rastrearse hasta el ancestro común de los deuterostomados. De hecho, las relaciones filogenéticas dentro de los clados de globina 1 y 3 podrían reflejar la presencia de genes de globina adicionales en el ancestro común de los

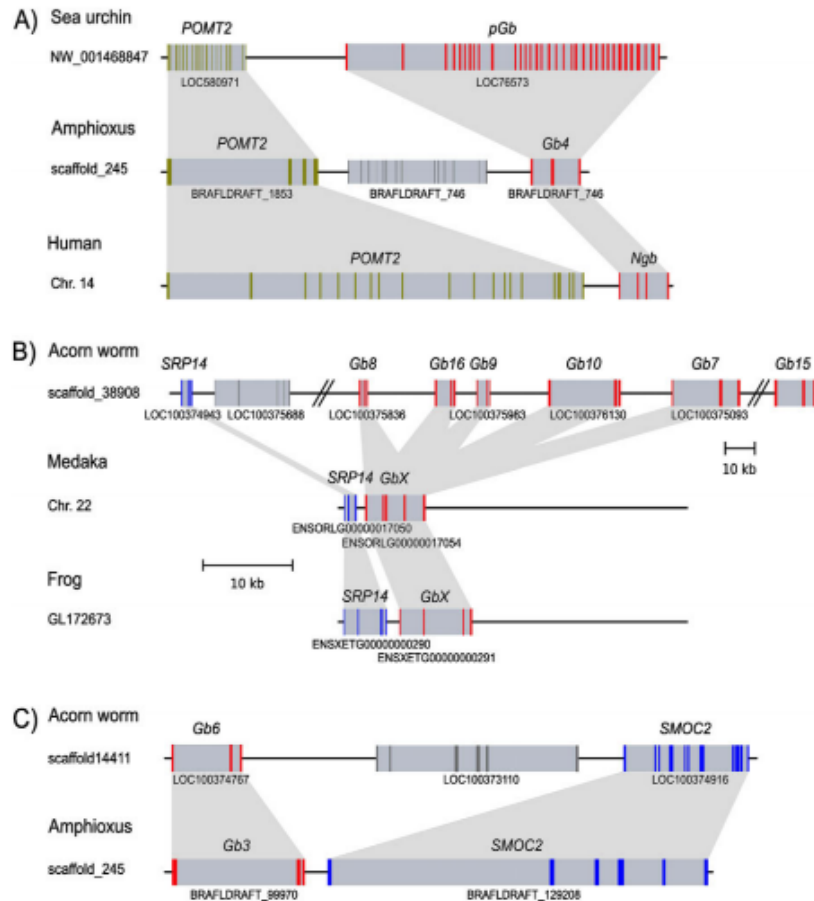


Fig. 3. Patterns of conserved synteny in the genomic regions that harbor paralogous genes from deuterostome globin clades 2–4. (A) Synteny comparisons of the *Ngb* region in acorn worm, amphioxus, and vertebrates (as represented by human). (B) Synteny comparisons of the *GbX* region in acorn worm, medaka, and frog. (C) Synteny comparisons between homologous chromosomal regions that contain the acorn worm *Gb6* and amphioxus *Gb3* genes. Nonglobin genes have been labeled according to their human ortholog. Intervening genes that do not contribute to conserved synteny are shown in gray. Horizontal lines denote orthologous relationships. The small scale applies to the acorn worm fragment that harbors four globin genes that are coorthologous to vertebrate *GbX*.

pueden ser más antiguos que la divergencia de Protostomia y Deuterostomia, y algunos de ellos pueden haberse originado antes de la división entre animales y plantas. Cabe señalar que este número inferido de genes de globina ancestrales representa una estimación mínima, ya que la adición de datos completos de la secuencia del genoma de deuterostomados adicionales podría dar lugar al descubrimiento

deuterostomados, y se puede hacer un caso similar para la secuencia divergente de amphioxus *Gb8* y para las globinas de gusano de bellota sin intrones.

Ganancias y pérdidas de genes de globina de deuterostomados

La distribución filética de los cuatro linajes de globina definidos y las relaciones filogenéticas

dentro de cada uno de los clados correspondientes revelan una historia compleja de duplicaciones y deleciones de genes específicos del linaje. Ninguno de los cinco filamentos de deuterostomados para los que tenemos datos de secuencia del genoma completo (Cephalochordata, Echinodermata, Hemichordata, Urochordata y Vertebrata) parece haber retenido representantes de las cuatro globinas ancestrales. Las especies incluidas en este estudio exhiben una tremenda variabilidad en el tamaño y composición de miembros de sus repertorios de genes de globina. Por ejemplo, los repertorios de globina del erizo de mar y las ascidias descienden de dos parálogos distintos que estaban presentes en el ancestro común del deuterostomado, y cada linaje genético parálogo se diversificó de forma independiente en los dos linajes organizmicos separados a través de repetidas rondas de duplicación y divergencia. Los limitados datos disponibles sugieren que las globinas del pepino de mar también derivan de varias rondas de duplicación y divergencia específicas del linaje. Sin embargo, hay varias subunidades de Hb de esta especie para las cuales no tenemos datos de secuencia (Kitto et al. 1998). Por lo tanto, se requerirán datos de secuencia genómica adicionales para obtener información más refinada sobre los orígenes de las globinas de pepino de mar.

A diferencia de los patrones observados en los erizos de mar y las ascidias, los representantes de hemicordados, cefalocordados y vertebrados han conservado copias ortólogas de al menos tres genes distintos que estaban presentes en el ancestro común del deuterostomado. Estos conjuntos de genes ancestrales se diversificaron independientemente en cada uno de los diferentes linajes organizmicos. Quizás el patrón más llamativo es el observado en el clado 4, que contiene Ngbs de vertebrados. En este caso, los ortólogos de Ngb se han perdido secundariamente en urocordados y hemicordados y se han retenido como genes de

copia única en cefalocordados y vertebrados. Sorprendentemente, el ortólogo de Ngb en el erizo de mar ha sufrido múltiples rondas de duplicación del dominio interno para que codifique un polipéptido con 16 dominios de globina unidos covalentemente (Bailly y Vinogradov 2008).

Evolución convergente de globinas de deuterostomados

Los análisis filogenéticos de las globinas de vertebrados han revelado los orígenes de nuevas propiedades funcionales y también han revelado casos en los que la evolución ha diseñado soluciones estructuralmente distintas a problemas fisiológicos similares (Berenbrink et al. 2005; Berenbrink 2007; Opazo et al. 2008; Hoffmann, Storz, et al. 2010; Hoffmann, Opazo, et al. 2010; Storz et al. 2011). Nuestros resultados demuestran que dentro de los deuterostomados, las globinas intracelulares circulantes con funciones de transporte de oxígeno evolucionaron tres veces independientemente en equinodermos, en vertebrados de ciclostomados y en vertebrados de gostostomados. Del mismo modo, el Ngb funcional de acoel, acoel Gb1, claramente no es ortólogo a los Ngbs de cordados, lo que indica que estas globinas nerviosas evolucionaron independientemente. De hecho, acoel Gb1 se coloca como hermana del pepino de mar que transporta oxígeno Hbs, en el clado de globina 1. Se ha informado un caso similar de reclutamiento de una globina de tipo Hb como globina nerviosa para el molusco *Spisula solidissima* (Dewilde et al. 2006). En el caso de las globinas circulantes de transporte de oxígeno, los mecanismos estructurales subyacentes de la unión cooperativa de oxígeno también evolucionaron de forma independiente. Las globinas no circulantes, como las de gnatostomados Cygb, GbE, GbX y las proteínas Ngb están presentes en los clados 1, 2 y 4, y las Hbs intracelulares circulantes con unión cooperativa de oxígeno se encuentran en tres

linajes diferentes dentro del clado 1 (en gnatostomados [vertebrados con mandíbula], ciclostromados [lampreas y pez bruja] y pepinos de mar). En la mayoría de los gnatóstomos, la cooperatividad de la Hb tetramérica $\alpha^2\beta^2$ proviene de un cambio ligado a la oxigenación en el enlace Fe^{2+} -F8 (His) que se propaga a los contactos $\alpha^1\beta^1$. La transición resultante en la estructura cuaternaria entre las conformaciones oxi y desoxi implica un reordenamiento de los enlaces de hidrógeno en los contactos entre subunidades formados por las hélices C, G y H (Perutz et al. 1987). Por el contrario, en los Hbs de transporte de oxígeno de los ciclostromados, la cooperatividad proviene de una disociación de homo y/o heterodímeros ligados a la oxigenación en monómeros de estado oxi (p. Ej., Brittain y Wells 1986; Brittain et al. 1989; Fago y Weber 1995 ; Qiu et al. 2000; Fago et al. 2001). Esta disociación dependiente de ligando está mediada por contactos entre subunidades que involucran las hélices E y F de tal manera que los grupos hemo están en contacto casi directo en el estado desoxi. Se observan modos similares de dinámica de asociación/disociación dependiente de ligando en los Hbs intracelulares y extracelulares de invertebrados (Kitto et al. 1998; Riggs 1998; Weber y Vinogradov 2001). Por ejemplo, en los Hbs intracelulares del pepino de mar, *Caudina arenicola*, la cooperatividad proviene de la asociación ligada a ligandos en tetrámeros desoxigenados y polímeros de nivel superior (Bonaventura y Kitto 1973; Mitchell et al.1995; Kitto et al.1998). Al igual que los Hbs de los ciclostromados y los de muchos taxones de invertebrados, los contactos entre subunidades en los glóbulos rojos Hbs de *Caudina* están formados por las hélices E y F. Sin embargo, las estructuras cristalinas de *Caudina* Hbs indican que los contactos entre subunidades involucran un conjunto único de residuos (Mitchell et al. 1995; Kitto et al. 1998).

Los resultados de nuestro análisis filogenético también arroja información sobre el mecanismo

involucrado en la evolución de las globinas multidominio. Mientras que Ngb es un monómero de dominio único en todos los cordados examinados hasta la fecha, el ortólogo de Ngb en el erizo de mar púrpura se convirtió en una proteína quimérica de 16 dominios a través de una combinación de fusión de genes y duplicación de dominios internos (Bailly y Vinogradov 2008). A medida que los proyectos de secuenciación del genoma revelan estructuras de genes de globina para una diversidad creciente de taxones de deuterostomados y protostomados, debería ser posible evaluar los roles relativos de la duplicación de dominios internos, la combinación de dominios y la fusión de genes en la evolución de las globinas multidominio complejas.