

# Aplicación de Metaheurísticas Multiobjetivo Basadas en Dominancia e Indicadores para Reconstrucción Filogenética

Sergio Santander-Jiménez y Miguel A. Vega-Rodríguez

Univ. Extremadura. Dept. Tecnología de los Computadores y de las Comunicaciones,  
Escuela Politécnica. Campus Universitario s/n, 10003. Cáceres, España  
{sesaji,mavega}@unex.es

**Resumen** La optimización de procesos biológicos que involucran varias funciones objetivo representa una de las líneas fundamentales de investigación en bioinformática. Dada la diversidad de estrategias multiobjetivo disponibles en la literatura, es preciso llevar a cabo estudios comparativos que permitan decidir qué diseño algorítmico conduce a los mejores resultados. En este trabajo abordamos la inferencia filogenética mediante la evaluación de dos líneas fundamentales de algoritmos multiobjetivo: aproximaciones basadas en dominancia (Non-Dominated Sorting Genetic Algorithm II y Strength Pareto Evolutionary Algorithm 2) y aproximaciones basadas en indicadores de calidad (Indicator-Based Evolutionary Algorithm y una nueva propuesta, el Indicator-Based Multiobjective Bat Algorithm). La experimentación llevada a cabo sobre cuatro bases de datos reales de nucleótidos verifica la bondad de combinar indicadores e inteligencia de enjambre, obteniéndose resultados significativos atendiendo al rendimiento multiobjetivo y biológico observado.

**Keywords:** Bioinformática, metaheurísticas multiobjetivo, filogenética

## 1. Introducción

El modelado de procesos biológicos como problemas de optimización tiene por objeto encontrar la solución más satisfactoria de acuerdo a un criterio de optimalidad que mida su calidad biológica. Los avances en desarrollos algorítmicos han propiciado la incorporación de supuestos más realistas en estos modelos, conduciendo a la necesidad de afrontar problemas que involucren múltiples criterios de manera simultánea. La reconstrucción filogenética representa uno de los problemas más relevantes en este contexto [6]. Dado un conjunto de secuencias moleculares alineadas, los procedimientos filogenéticos tratan de explicar las características observables en los organismos caracterizados mediante la inferencia de su historia evolutiva. Una de las razones que fundamentan su formulación desde una perspectiva multiobjetivo es la necesidad de afrontar los problemas de incongruencia que se observan durante el proceso de inferencia, como ocurre con la definición de distintos criterios de optimalidad. Coelho et al. afrontaron

S. Santander-Jiménez y M. A. Vega-Rodríguez

dichos problemas mediante una aproximación inmuno-inspirada según los criterios de error cuadrático medio y evolución mínima [2], mientras que Cancino y Delbem sugirieron el uso de optimización multiobjetivo para realizar análisis filogenéticos según los populares principios de parsimonia y verosimilitud [1].

Este trabajo trata de ir un paso más allá en el desarrollo de soluciones multiobjetivo para afrontar problemas reales de alta complejidad como es el caso de la inferencia filogenética. Por ello, proponemos un estudio comparativo entre diferentes diseños multiobjetivo de acuerdo a dos líneas básicas de desarrollo: métodos basados en dominancia y métodos basados en indicadores de calidad. En concreto, consideramos cuatro metaheurísticas: Non-Dominated Sorting Genetic Algorithm II (NSGA-II) [4], Strength Pareto Evolutionary Algorithm 2 (SPEA2) [12] y el Indicator-Based Evolutionary Algorithm (IBEA) [11], así como un nuevo diseño de inteligencia de enjambre basado en el Bat Algorithm [10], el Indicator-Based Multiobjective Bat Algorithm (IMOA). El análisis de rendimiento de estas aproximaciones se efectuará sobre cuatro bases de datos de nucleótidos, empleando diversas métricas multiobjetivo para decidir qué algoritmo presenta el mejor comportamiento a la hora de resolver el problema.

Este artículo se organiza del siguiente modo. La Sección 2 describe la formulación multiobjetivo del problema. La Sección 3 resume las características de los diseños multiobjetivo evaluados, cuyos resultados son presentados en la Sección 4. Finalmente, la Sección 5 detalla conclusiones y líneas futuras de trabajo.

## 2. Reconstrucción Filogenética

La inferencia filogenética trata de proporcionar hipótesis sobre la evolución de los organismos vivos en base a las divergencias y similitudes observadas en sus secuencias moleculares. La entrada de un procedimiento filogenético viene dada por una matriz de caracteres de dimensiones  $N \times M$  que contiene, en el caso de análisis basados en ADN, secuencias alineadas de nucleótidos correspondientes a las especies de las cuales se pretende inferir su historia evolutiva (donde  $N$  es el número de especies y  $M$  la longitud de sus secuencias). Sus relaciones evolutivas son modeladas mediante topologías filogenéticas  $T=(V, E)$ , donde  $V$  es el conjunto de nodos que contiene los organismos caracterizados en la entrada (hojas), así como organismos ancestrales hipotéticos (nodos internos), y  $E$  es el conjunto de ramas que establecen relaciones ancestro-descendiente entre nodos.

Al modelar la reconstrucción filogenética como un problema de optimización, es necesario precisar que la elección del criterio de optimalidad representa una de las fuentes de incongruencia más significativas en escenarios reales. En este trabajo consideramos la resolución del problema conforme a los principios de parsimonia y verosimilitud. Por un lado, la parsimonia tiene por objeto inferir el árbol filogenético que minimice la cantidad de cambios evolutivos observados entre especies emparentadas. Por su parte, la verosimilitud asume los supuestos dado por un modelo evolutivo  $\Phi$  para inferir la historia evolutiva más plausible, midiendo la probabilidad condicional de que una determinada hipótesis filogenética haya dado lugar a las secuencias observadas en la entrada.

### Metaheurísticas Multiobjetivo para Reconstrucción Filogenética

Sea  $S$  el espacio de posibles topologías filogenéticas. La formulación multiobjetivo del problema busca encontrar aquellos árboles Pareto  $T \in S$  que optimicen simultáneamente las funciones de parsimonia  $P(T)$  y verosimilitud  $L(T, \Phi)$ :

$$\begin{aligned} \text{optimizar } \vec{f}(T) &= \{f_1(T), f_2(T)\}, \\ \text{donde } f_1(T) &= \text{minimizar } P(T), \\ f_2(T) &= \text{maximizar } L(T, \Phi). \end{aligned} \tag{1}$$

El cálculo de  $P(T)$  para una solución candidata  $T=(V, E)$  sobre un conjunto de datos de tamaño  $N \times M$  se puede realizar conforme a lo siguiente [6]:

$$P(T) = \sum_{i=1}^M \sum_{(a,b) \in E} C(a_i, b_i), \tag{2}$$

donde  $(a, b) \in E$  hace referencia a la relación evolutiva entre dos nodos  $a, b \in V$ ,  $a_i$  y  $b_i$  son los estados observados en el carácter  $i$ -ésimo de las secuencias asociadas a  $a$  y  $b$ , y  $C(a_i, b_i)$  cuantifica la presencia (1) o no (0) de un cambio de estado entre  $a_i$  y  $b_i$ .

Por su parte, el cálculo de  $L(T, \Phi)$  sigue la siguiente formulación [6]:

$$L[T, \Phi] = \prod_{i=1}^M \sum_{x, y \in A} \pi_x [P_{xy}(t_{ru}) L_p(u_i = y)] \times [P_{xy}(t_{rv}) L_p(v_i = y)], \tag{3}$$

donde  $A$  es el alfabeto de estados de carácter,  $\pi_x$  la probabilidad estacionaria del estado  $x \in A$ ,  $P_{xy}(t)$  la probabilidad de mutación desde  $x$  a otro estado  $y$  en un tiempo  $t$ ,  $r \in V$  la raíz de  $T$  con descendientes  $u, v \in V$  y  $L_p(u_i = y)$ ,  $L_p(v_i = y)$  la verosimilitud parcial de observar  $y$  en el carácter  $i$ -ésimo de  $u$  y  $v$ .

## 3. Aproximaciones Multiobjetivo

Este estudio evalúa el rendimiento mostrado por dos líneas principales de algoritmos multiobjetivo a la hora de realizar análisis filogenéticos. Para adaptar estos algoritmos al problema, se ha empleado una representación del individuo introducida en [7] y que se fundamenta en el uso de matrices  $m$  de distancias genéticas. La idea fundamental consiste en implementar los operadores evolutivos en base a estas matrices y la aplicación del método BIONJ [6] para inferir las topologías filogenéticas correspondientes. Dado que empleamos algoritmos basados en población, la inicialización de la misma se realiza considerando un repositorio de soluciones iniciales generadas mediante técnicas de *bootstrapping*.

### 3.1. Métodos Basados en Dominancia

Estos diseños algorítmicos dirigen la búsqueda distinguiendo la calidad de las soluciones conforme al concepto de dominancia Pareto [3]. Así, el proceso de

S. Santander-Jiménez y M. A. Vega-Rodríguez

optimización tiene por objeto minimizar la distancia al frente de Pareto óptimo y maximizar la diversidad de las soluciones [12], incorporando para ello mecanismos de clasificación Pareto y estimación de densidad. Dentro de esta línea, consideramos dos algoritmos evolutivos multiobjetivo: NSGA-II y SPEA2.

**Non-Dominated Sorting Genetic Algorithm II** NSGA-II [4] considera el manejo de una población padre  $P$  y una población *offspring*  $Q$ . En cada generación, se combinan ambas poblaciones en una población auxiliar  $C$  que es procesada mediante los métodos de ordenación rápida no dominada y el cálculo de distancias de *crowding*. Los mejores individuos en  $C$  son posteriormente usados para actualizar  $P$ , aplicándose sobre la misma operadores genéticos para generar nuevas soluciones en  $Q$ . Nuestra implementación del algoritmo para filogenética incluye los siguientes operadores. Como operador de selección, se aplican torneos binarios basados en los valores de *ranking* y *crowding* obtenidos por cada solución. La generación de nuevas soluciones se efectúa mediante un operador de cruce uniforme que intercambia filas completas de las matrices padre para generar una nueva matriz hija. Finalmente, el operador de mutación modifica entradas aleatorias en la matriz hija según una distribución gamma.

**Strength Pareto Evolutionary Algorithm 2** SPEA2 [12] maneja una población de individuos  $P$  y un archivo  $\bar{P}$  que contiene una muestra representativa de las mejores soluciones encontradas. En cada generación, se computa el fitness de cada individuo  $s$  en  $P$  y  $\bar{P}$  considerando los valores de *strength* de los individuos que dominan a  $s$ , junto con un valor de densidad dado por la inversa de la distancia al  $k$ -ésimo vecino más próximo. Tras ello, se procede con la actualización de  $\bar{P}$ , copiando los individuos no dominados de  $P$  y ajustando al tamaño del archivo mediante procedimientos de relleno o truncado. El archivo actualizado es usado posteriormente con propósitos de selección durante la generación de nuevas soluciones hijas. La adaptación de SPEA2 al problema considera los mismos operadores de cruce y mutación indicados para NSGA-II, así como selección por torneo binario basado en los valores de fitness definidos por el algoritmo.

### 3.2. Métodos Basados en Indicadores

La segunda clase de algoritmos estudiados son los diseños basados en indicadores, que integran métricas de rendimiento multiobjetivo dentro del cálculo de fitness y selección de soluciones. Por tanto, el proceso de optimización pretende generar el conjunto de soluciones que optimice el indicador elegido. Nuestro estudio incluye como representantes de esta línea a IBEA e IMOBA. En nuestras implementaciones, se ha aplicado el indicador  $I_{HD}$  [11], basado en la métrica de hipervolumen  $I_H$ , que compara dos conjuntos de soluciones  $X$ ,  $Y$  mediante el cómputo del área/volumen del espacio objetivo dominado por  $Y$  pero no por  $X$ .

**Indicator-Based Evolutionary Algorithm** IBEA [11] emplea una población  $P$  donde se distingue la calidad de sus individuos mediante comparaciones por

Metaheurísticas Multiobjetivo para Reconstrucción Filogenética

pares según el indicador de calidad elegido. Concretamente, el cálculo de fitness para un individuo  $s$  viene dado por  $s.Fitness = \sum_{s' \in P \setminus \{s\}} -e^{-I_{HD}(\{s'\}, \{s\})/c\kappa}$ , donde  $c$  es el valor absoluto máximo del indicador y  $\kappa$  un factor de escala. En cada generación, se aplica un procedimiento de reajuste de la población donde se identifica el individuo  $s_w$  que presente el peor valor de fitness. Dicho individuo es eliminado de  $P$ , actualizándose los fitness de los individuos supervivientes como  $s.Fitness = s.Fitness + e^{-I_{HD}(\{s_w\}, \{s\})/c\kappa}$ . Este procedimiento se repite hasta que el tamaño de  $P$  se adecúa al tamaño de la población fijado como parámetro de entrada. Tras ello, la población resultante es empleada para generar nuevas soluciones candidatas hijas. Se ha aplicado la versión adaptativa de IBEA al problema, incluyendo selección por torneo binario basado en los valores de fitness de IBEA junto con los operadores de cruce y mutación descritos previamente.

**Indicator-Based Multiobjective Bat Algorithm** IMOBA es una adaptación multiobjetivo del Bat Algorithm [10]. Al comienzo, cada murciélago en  $P$  es inicializado asignando una posición inicial dada por una matriz de distancias  $m_0$ , una matriz de velocidades  $v_0 = 0$ , una frecuencia de pulso sonora  $f_0$  en el rango  $[f_{\min}, f_{\max}]$ , un ratio de emisión de pulsos  $r_0 \in [0, r_{\max}]$  y una intensidad de onda  $A_0$ . En cada generación, se produce la selección y reajuste de  $P$  para mantener los  $tamPop$  murciélagos más prometedores, considerando el método de asignación de fitness de IBEA. A partir de esta selección se modela la primera sección del algoritmo: el paso de movimiento (líneas 5-8 en el Pseudocódigo 1). Sea  $s$  un murciélago con velocidad  $s.v$  y frecuencia  $s.f$  cuya solución viene dada por  $s.m$ . Se calcula una nueva posición para  $s$  usando:

$$\begin{aligned} s.f &= f_{\min} + (f_{\max} - f_{\min})\beta, \\ s.v[i, j] &= s.v[i, j] + (s.m[i, j] - s_{rng}.m[i, j])s.f, \\ s.m[i, j] &= s.m[i, j] + s.v[i, j], \end{aligned} \tag{4}$$

donde  $\beta \in [0, 1]$  es un número aleatorio de una distribución uniforme y  $s_{rng}$  un murciélago elegido aleatoriamente de  $P$  que actúa como referencia en el movimiento de  $s$ . Las soluciones obtenidas son integradas en la población, procediéndose de nuevo a su reajuste para mantener los mejores  $tamPop$  murciélagos.

Posteriormente, se calcula un array de probabilidades de selección de cada solución según sus valores de fitness, empleándose en la siguiente sección. El paso de búsqueda (líneas 13-20) involucra un proceso de explotación-exploración cuya decisión depende del ratio actual de emisión  $s.r$  de cada murciélago (línea 14, donde  $rand$  es un número aleatorio en el rango  $[0, 1]$ ). En la rama de explotación, se genera una nueva solución  $s_{new}$  aplicando la expresión  $s_{new}.m[i, j] = s_{sel}.m[i, j] + \varepsilon(A)$ , donde  $s_{sel}$  se refiere a un murciélago en  $P$  seleccionado usando el array de probabilidades,  $\varepsilon$  es un número aleatorio en el intervalo  $[-1, 1]$ , y  $(A)$  es la intensidad promedio observada en  $P$ . Por su parte, la rama de exploración involucra la generación de  $s_{new}$  mediante la selección de una nueva solución de partida del repositorio, aplicando optimización topológica SPR y NNI [6].

Los pasos finales de cada generación contemplan la actualización de los valores de intensidad  $s.A$  y ratios de emisión de pulsos  $s.r$  de cada murciélago  $s \in P$ .

S. Santander-Jiménez y M. A. Vega-Rodríguez

---

**Pseudocódigo 1** Indicator-Based Multiobjective Bat Algorithm

---

```

1: FrentePareto  $\leftarrow 0$ ,  $P \leftarrow$  Inicializar Población ( $tamPop$ ,  $f_{m\acute{a}x}$ ,  $f_{m\acute{i}n}$ ,  $A_0$ ,  $r_{max}$ )
2: mientras ! criterio de parada ( $maxEval$ ) hacer
3:    $P_i \leftarrow$  Calcular Fitness Basado en Indicador ( $P$ ,  $\kappa$ ,  $Z_{ref}$ ) /*  $\forall i: i=1$  to  $|P|$  */
4:    $P \leftarrow$  Selección ( $tamPop$ ,  $\kappa$ ) /* Salvando los mejores  $tamPop$  murciélagos */
5:   para  $i = 1$  to  $tamPop$  hacer
6:      $P'_i.m \leftarrow$  Mover Murciélago ( $P_i$ ,  $P$ ,  $f_{m\acute{a}x}$ ,  $f_{m\acute{i}n}$ )
7:      $P'_i.T \leftarrow$  Inferir y Evaluar Solución ( $P'_i.m$ )
8:   fin para
9:    $P \leftarrow P \cup P'$ 
10:   $P_i \leftarrow$  Calcular Fitness Basado en Indicador ( $P$ ,  $\kappa$ ,  $Z_{ref}$ ) /*  $\forall i: i=1$  to  $|P|$  */
11:   $P \leftarrow$  Selección ( $tamPop$ ,  $\kappa$ ) /* Salvando los mejores  $tamPop$  murciélagos */
12:  arraySelección  $\leftarrow$  Calcular Probabilidades de Selección ( $P$ ,  $tamPop$ )
13:  para  $i = 1$  to  $tamPop$  hacer
14:    si  $rand > P_{i,r}$  entonces
15:       $P'_i.m \leftarrow$  Generar Solución - Explotación ( $P$ , arraySelección,  $probMutacion$ )
16:    si no
17:       $P'_i.m \leftarrow$  Generar Solución - Exploración ( $probMutacion$ )
18:    fin si
19:     $P'_i.T \leftarrow$  Inferir y Evaluar Solución ( $P'_i.m$ )
20:  fin para
21:   $P \leftarrow P \cup P'$ 
22:   $P_i \leftarrow$  Actualizar Intensidad y Ratios Emisión ( $\alpha$ ,  $P_i.A$ ,  $\gamma$ ,  $P_i.r_{m\acute{a}x}$ ) /*  $\forall i: i=1$  to  $|P|$  */
23:  FrentePareto  $\leftarrow$  Actualizar Frente Pareto ( $P$ , FrentePareto)
24: fin mientras
    
```

---

Para ello, se definen las expresiones  $s.A = \alpha s.A$  y  $s.r = s.r_{m\acute{a}x}[1 - \exp(-\gamma t)]$ , donde  $\alpha$  representa el factor de actualización de la intensidad de onda,  $\gamma$  el factor de actualización de emisión y  $t$  identifica la generación actual.

#### 4. Resultados Experimentales

Con objeto de examinar el rendimiento obtenido por estas metaheurísticas, se ha realizado experimentación sobre cuatro bases de datos reales de nucleótidos [1,7]: rbcL55, mtDNA.186, RDPII.218 y ZILLA.500. Cada experimento involucra 31 ejecuciones independientes para cada algoritmo, realizadas sobre una estructura hardware compuesta por un procesador AMD Opteron Magny-Cours 6174 con 16GB DDR3 RAM y Scientific Linux 6.1 como sistema operativo. El software evaluado ha sido compilado con GCC 4.4.5 (usando la bandera -O3).

La Tabla 1 muestra los resultados de los estudios paramétricos efectuados para configurar los algoritmos bajo estudio. Nótese que la intensidad inicial  $A_0$  en IMOBA toma como valor el parámetro de forma de la distribución gamma observada en las distancias genéticas para cada base de datos [6].

**Evaluación de Resultados** A fin de verificar la calidad de los frentes de Pareto obtenidos, se han empleado 3 métricas multiobjetivo [3]: el *hipervolumen*  $I_H$  del espacio objetivo cubierto por cada algoritmo, la *relación de cobertura*  $SC(X, Y)$  que proporciona la fracción de soluciones de  $Y$  que son débilmente dominadas por  $X$  y el indicador de *espaciado*  $SP$  para evaluar la uniformidad de la distribución de soluciones observada en los frentes. Las Tablas 2 y 3 detallan los resultados medianos obtenidos por cada algoritmo bajo estas tres métricas. Así mismo, la



S. Santander-Jiménez y M. A. Vega-Rodríguez

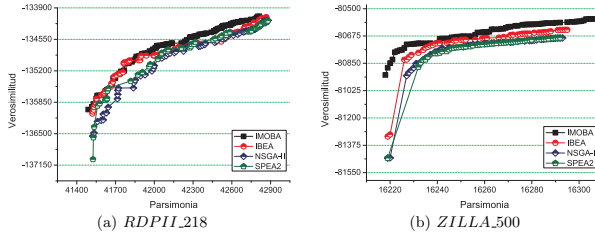


Figura 1. Evaluación de rendimiento multiobjetivo - Frentes de Pareto

Con respecto al espaciado, los métodos basados en indicadores muestran también un rendimiento significativo atendiendo a la distribución de soluciones en los frentes de Pareto. En este sentido, IMOBA obtiene el mejor rendimiento en términos generales (valor medio de 0,04 frente al 0,05 de IBEA). Finalmente, la comparativa por pares de los frentes bajo la relación de cobertura da soporte a las ideas sugeridas por las otras métricas, apuntando a IMOBA como la aproximación más exitosa. Concretamente, este algoritmo cubre porcentajes entre 75,29 % y 100,00 % de las soluciones obtenidas por IBEA, dominando a su vez a NSGA-II y SPEA2 en todas las instancias (98,25 % - 100,00 %).

Estos resultados sugieren que la aplicación de metaheurísticas basadas en indicadores en filogenética permite mejorar la búsqueda de árboles Pareto de gran calidad, empleando la información proporcionada por el indicador considerado para guiar el mecanismo de búsqueda. Además, los principios de inteligencia de enjambre permiten a IMOBA afrontar el problema mediante la definición de diversos mecanismos (movimiento, explotación y exploración) que consideran la información obtenida por la población completa de murciélagos para mejorar las capacidades de búsqueda. Como resultado, esta propuesta proporciona una mejora significativa en términos de calidad de resultados, manteniendo a su vez los tiempos de ejecución en el mismo orden que el resto de algoritmos.

**Comparaciones con Otras Herramientas** A continuación, comparamos la propuesta IMOBA con otros métodos multiobjetivo para reconstrucción filogenética: una aproximación previa de inteligencia de enjambre basado en dominancia Multiobjective Artificial Bee Colony (MOABC) [7] y la herramienta PhyloMOEA de Cancino y Delbem [1]. La Tabla 4 presenta los resultados de MOABC e IMOBA atendiendo a las tres métricas multiobjetivo empleadas en este trabajo. La comparativa da cuenta del rendimiento significativo alcanzado por IMOBA con respecto a MOABC, verificando una mejora estadísticamente significativa en términos de hipervolumen en las instancias mtDNA.186, RD-P11.218 y ZILLA.500. Si bien ambos algoritmos obtienen resultados similares



Metaheurísticas Multiobjetivo para Reconstrucción Filogenética

**Tabla 4.** Comparativa multiobjetivo con MOABC

Instancia	Hipervolumen		Espaciado	
	$I_H$ (MOBA)	$I_H$ (MOABC)	SP(IMOBA)	SP(MOABC)
rbcl.55	71.60%±0.02	<b>71.73%±0.02</b>	0.084	<b>0.071</b>
mtDNA.186	<b>70.03%±0.01</b>	70.02%±0.01	0.055	<b>0.041</b>
RDPII.218	<b>75.06%±0.04</b>	74.64%±0.05	<b>0.012</b>	0.013
ZILLA.500	<b>73.08%±0.01</b>	73.02%±0.01	<b>0.013</b>	0.024
Relación de cobertura				
Instancia	SC(IMOBA, MOABC)		SC(MOABC, IMOBA)	
rbcl.55	20.00%		<b>71.43%</b>	
mtDNA.186	<b>89.47%</b>		13.33%	
RDPII.218	<b>93.88%</b>		1.60%	
ZILLA.500	<b>93.10%</b>		0.00%	

**Tabla 5.** Comparativa con PhyloMOEA (HKY85+ $\Gamma$ )

Instancia	Mejor parsimonia	
	IMOBA	PhyloMOEA
rbcl.55	<b>4874</b>	<b>4874</b>
mtDNA.186	<b>2431</b>	2437
RDPII.218	<b>41488</b>	41534
ZILLA.500	<b>16218</b>	16219
Mejor verosimilitud		
rbcl.55	<b>-21812.4</b>	-21889.8
mtDNA.186	<b>-39888.4</b>	-39896.4
RDPII.218	<b>-134151.1</b>	-134696.5
ZILLA.500	<b>-80965.7</b>	-81018.1

**Tabla 6.** Comparativas de parsimonia y verosimilitud (GTR+ $\Gamma$ ) con TNT y RAxML

Instancia	Parsimonia		Verosimilitud	
	IMOBA	TNT	IMOBA	RAxML
rbcl.55	<b>4874</b>	<b>4874</b>	<b>-21781.74</b>	-21788.12
mtDNA.186	<b>2431</b>	<b>2431</b>	<b>-39868.61</b>	-39868.91
RDPII.218	<b>41488</b>	<b>41488</b>	<b>-134070.98</b>	-134096.96
ZILLA.500	<b>16218</b>	<b>16218</b>	<b>-80566.19</b>	-80598.90

atendiendo a la métrica de espaciado, la relación de cobertura indica que un porcentaje notable de las soluciones de MOABC son dominadas por IMOBA. Con respecto a PhyloMOEA, la Tabla 5 muestra que IMOBA obtiene, en la ejecución de hipervolumen mediano, árboles de calidad atendiendo a sus valores de parsimonia y verosimilitud, mejorando los resultados publicados en [1].

Inciendiendo en la evaluación de resultados biológicos, la Tabla 6 muestra una comparativa con las herramientas TNT [5] (parsimonia) y RAxML [9] (verosimilitud). De acuerdo a los resultados obtenidos, IMOBA es capaz de igualar los valores de máxima parsimonia de TNT en las cuatro instancias analizadas, así como de inferir árboles de máxima verosimilitud de calidad significativa con respecto a RAxML. En resumen, estos resultados sugieren la bondad de emplear aproximaciones que integren indicadores e inteligencia de enjambre para inferir soluciones de gran relevancia atendiendo a su calidad multiobjetivo y biológica.

**5. Conclusiones**

Este trabajo ha profundizado en el estudio de metaheurísticas multiobjetivo para filogenética. El principal objetivo radicaba en la evaluación de dos de las líneas de desarrollo más populares en optimización multiobjetivo: aproximaciones basadas en dominancia (representadas por NSGA-II y SPEA2) y aproximaciones basadas en indicadores de calidad (IBEA e IMOBA, una nueva propuesta inspirada en las propiedades de ecolocalización de los murciélagos). Nuestra experimentación ha dado cuenta de la relevancia de aplicar propuestas basadas en indicadores a este problema, mejorando a los algoritmos basados en dominancia atendiendo a diversas métricas multiobjetivo. En este sentido, IMOBA representa un paso más allá con respecto al resto de algoritmos, posicionándose como un método robusto según las comparativas efectuadas con otras propuestas del

S. Santander-Jiménez y M. A. Vega-Rodríguez

estado del arte. Por tanto, la combinación de búsquedas basadas en indicadores e inteligencia de enjambre representa una aproximación prometedora para abordar problemas de gran complejidad como la inferencia filogenética.

Como trabajo futuro se pretende estudiar la integración de metaheurísticas y paralelismo de grano fino con GPUs, así como técnicas de supercomputación, para minimizar los tiempos de inferencia. A su vez, se estudiarán equipos paralelos de metaheurísticas para mejorar la calidad de los resultados. Otras líneas de trabajo incluyen la evaluación de IMOBA en escenarios benchmark, comparativas con algoritmos de descomposición y su aplicación en otros problemas.

**Agradecimientos** Los autores agradecen a la Universidad de Extremadura el apoyo económico ofrecido a Sergio Santander-Jiménez dentro del Plan de Iniciación a la Investigación, Desarrollo Tecnológico e Innovación 2015 (ACCION-III-04). Gracias a la Junta de Extremadura por la ayuda GR15011 concedida al grupo de investigación TIC015.

## Referencias

1. Cancino, W., Delbem, A.C.B.: A Multi-Criterion Evolutionary Approach Applied to Phylogenetic Reconstruction. En: *New Achievements in Evol. Comp.*, pp. 135–156. InTech (2010)
2. Coelho, G.P., Silva, A.E.A., Zuben, F.J.V.: An Immune-Inspired Multi-Objective Approach to the Reconstruction of Phylogenetic Trees. *Neural Comput. Appl.* 19(8), 1103–1132 (2010)
3. Coello, C., Dhaenens, C., Jourdan, L.: *Advances in Multi-Objective Nature Inspired Computing*. Springer Verlag, Berlin / Heidelberg (2010)
4. Deb, K., Pratap, A., Agarwal, S., Meyarivan, T.: A Fast and Elitist Multi-Objective Genetic Algorithm: NSGA-II. *IEEE Trans. Evol. Comput.* 6(2), 182–197 (2002)
5. Goloboff, P.A., Farris, J.S., Nixon, K.C.: TNT, a free program for phylogenetic analysis. *Cladistics* 24(5), 774–786 (2008)
6. Lemey, P., Salemi, M., Vandamme, A.M.: *The Phylogenetic Handbook: a Practical Approach to Phylogenetic Analysis and Hypothesis Testing*. Cambridge Univ. Press, Cambridge (2009)
7. Santander-Jiménez, S., Vega-Rodríguez, M.A.: Análisis Comparativo de Implementaciones del Algoritmo Multiobjective Artificial Bee Colony para Reconstrucción Filogenética. En: *Actas del X Congreso Español sobre Metaheurísticas, Algoritmos Evolutivos y Bioinspirados*, Mérida, Spain, pp. 369–376 (2015)
8. Sheskin, D.J.: *Handbook of Parametric and Nonparametric Statistical Procedures*. 5th edition. Chapman & Hall/CRC Press, NY, USA (2011)
9. Stamatakis, A.: RAxML Version 8: A Tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. *Bioinformatics* 30(9), 1312–1313 (2014)
10. Yang, X.S.: Bat Algorithm: Literature Review and Applications. *Int. J. Bio-Inspired Computation* 5(3), 141–149 (2013)
11. Zitzler, E., Künzli, S.: Indicator-Based Selection in Multiobjective Search. En: *PPSN VIII. LNCS*, vol. 3242, pp. 832–842. Springer Verlag (2004)
12. Zitzler, E., Laumanns, M., Thiele, L.: SPEA2: Improving the Strength Pareto Evolutionary Algorithm. En: *Proc. of EUROGEN'02*, pp. 95–100 (2002)